AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 16, NUMBER 5 MAY, 1948

Physical Aspects of Bacterial Growth and Multiplication

ALLEN L. KING
Dartmouth College, Hanover, New Hampshire

THE problem of bacterial growth and multiplication is especially intriguing to the physicist and mathematician. Unlike many biological problems, this one may be investigated fairly easily by physical and mathematical methods. Even so, the theoretical results have not always been in agreement with experimental observation, and a good explanation for some observations is entirely lacking. A well defined quantitative theory of bacterial growth and multiplication, which attempts to bring together into one framework a wide variety of experimental data and unexplained observations, should be built on fundamental principles. It is not surprising to find these principles largely physical in character. Although the present discussion is restricted to bacteria, the analysis may be modified for other unicellular organisms.

As is the common practice in physics, a finite number of assumptions are made. In the first place, the growth and multiplication of a bacterium is assumed to take place in a suitable liquid medium at a uniform temperature, under constant pressure and in darkness. At any instant the environment of each cell is similar to that of every other cell in the medium; the medium is infinite in extent. For simplicity, the medium is assumed to contain only one critical substance, at the moment undetermined, which affects the rate of growth and the limit of the multiplication process. The number of molecules of this substance required by an organism to double its volume is constant for all generations and is independent of temperature. The medium is inoculated with actively growing organisms essentially in the same phase of growth. The physiological characteristics of the cell-medium interface is assumed to remain unchanged from generation to generation.[1] No deaths occur; death, or permanent cessation of fission-ability, probably is relatively infrequent until the generation time becomes very great.[2] Finally, it is assumed that both the bacteria and the food particles are subject to random thermal motions, so that elementary kinetic theory may be applied.

Growth of a single cell.—The number of molecules of the critical substance striking unit area per unit time, because of random thermal motion, is given by the well-known relation from kinetic theory, $\frac{1}{4}n\bar{c}$, where n is the number of these molecules per unit volume and $\bar{c}$ is their average speed. A bacterium will meet $\frac{1}{4}n\bar{c}A$ of these molecules in unit time, where A is the area of the bacterium surface at a given instant. Only a fraction α of the molecules is assumed to have sufficient energy to penetrate the cell wall and become assimilated by the bacterium. Let b represent the number of these molecules which must be assimilated, regardless of the post-fission volume V_0, before a cell can split by fission. Then the number of critical molecules required, on the average, per unit volume of cell is b/V_0,

[1] M. A. Barber, *J. Infectious Diseases* **5**, 379 (1908).
[2] M. W. Jennison, *J. Bact.* **33**, 461 (1937).

and the rate of increase of volume of a single cell may be written $dV/dt = \frac{1}{4}\alpha\bar{c}V_0An/b$. Meanwhile the rate of decrease of the critical substance is $dn/dt = -\frac{1}{4}\alpha\bar{c}ANn$, where N is the number of bacteria per unit volume during a given generation. At the start of a generation the volume of a cell is V_0 and the number of critical molecules equals $n_0 - b(N - N_0)$, where n_0 and N_0 are the initial values of n and N, respectively. Elimination of n between these equations yields the differential equation of growth of a single cell, namely,

$$dV/dt = (n_0 + bN_0 - bNV/V_0)\alpha\bar{c}V_0A/4b. \quad (1)$$

Fortunately, for the case of rod-like bacteria of constant radius a, Eq. (1) can be integrated fairly easily. For the first few generations ($bN \ll n_0$) the result is

$$\ln(l/l_0) = (\alpha\bar{c}n_0V_0/2ba)t. \quad (2)$$

That the logarithm of the length of rod-like bacteria is a linear function of time has been shown by Schmalhausen and Bordzilowskaja[3] for *Escherichia coli*, *Proteus vulgaris* and *Shigella dysenteriae;* by Bayne-Jones and Adolph[4] for *Bacillus megatherium;* by Adolph[5] for *Bacillus mycoides;* and by Knaysi[6] for *Bacillus cereus*. Rahn[7] has discussed this problem in some detail.

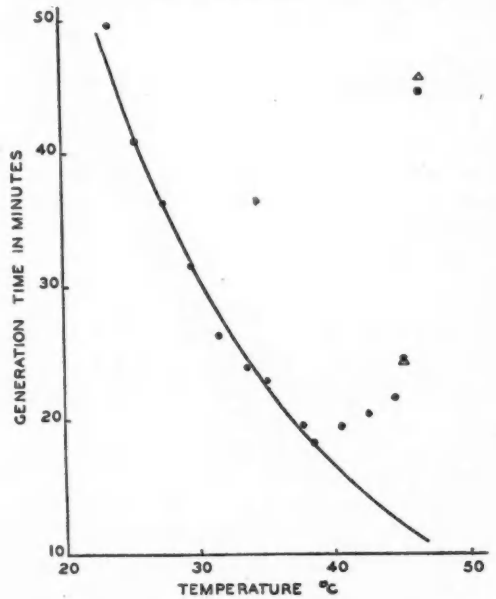

FIG. 1. Effect of temperature on the generation time of *Escherichia coli*. The dots are plotted from Barber's data. The curve is the graphical representation of Eq. (4). The triangles indicate two points also evaluated by means of Eq. (4).

[3] I. Schmalhausen and N. Bordzilowskaja, *Z. wiss. Biol.* **107**, 672 (1926).

Effect of temperature on generation time.—The time required for a bacterium to double its volume and divide, namely, the "generation time τ," may be measured as a function of temperature. For instance, Barber[1] carried out a long series of careful measurements on *Escherichia coli* several years ago. I have selected all his measurements where there were at least eight determinations of generation time at a given temperature. Usually each determination was made by observing the time for the single cell to multiply three to five times. These data are shown as dots in Fig. 1. Note that the generation time reaches a minimum value near 40°C. Thereafter, the points are scattered with a rapid increase in generation time. This scattering is not a sign of death, for the data were obtained on living, growing cells. Barber found that multiplication ceased beyond 49°C. He further observed that below 40°C the cells were in their long form, but at temperatures above 45°C they were short, being similar in appearance to cells in an old culture. Between 40° and 45°C there was a mixture of the two forms. These observations have an important bearing on the interpretation of the divergence between the measured and theoretical values of generation time for temperatures above 40°C.

A relation for the generation time of rod-like bacteria during the first few generations may be found by setting l equal to $2l_0$ in Eq. (2); thus

$$\tau = 2ba \ln 2/\alpha\bar{c}n_0V_0. \quad (3)$$

Over a wide range of temperatures, τ ordinarily depends on temperature through α and $\bar{c}$ only. From kinetic theory the average speed $\bar{c}$ of the molecules is equal to $(8kT/\pi m)^{\frac{1}{2}}$, in which k is the Boltzmann constant, T is the absolute temperature of the medium, and m is the mass of a molecule of the critical substance. If all the

[4] S. Bayne-Jones and E. F. Adolph, *Cellular and Comparative Physiol.* **1**, 409 (1932).

[5] E. F. Adolph, *The regulation of size as illustrated in unicellular organisms* (Thomas, Springfield, Ill., 1931).

[6] G. Knaysi, *J. Bact.* **40**, 247 (1940).

[7] O. Rahn, *Mathematics in bacteriology* (Burgess, Minneapolis, 1939), pp. 13–16.

molecules that strike the surface of a bacterium with energies exceeding an optimum value w are captured and assimilated by the organism, then, to a first approximation,

$$\alpha = \tfrac{1}{2}(kT/w)^{\frac{1}{2}} \exp(-w/kT).$$

The energy w appears to be a fundamental constant characterizing a unicellular organism. Substitution of these expressions for α and $\bar{c}$ in Eq. (3) yields

$$\tau = (B/T) \exp(T^*/T), \qquad (4)$$

where T^* equals w/k, and B is the constant $0.693ba(2\pi mw)^{\frac{1}{2}}/kn_0V_0$. The curve of Fig. 1 is a graph of Eq. (4) with T^* set equal to 5560°K and B set equal to 1.01×10^{-4} min. deg. Computed generation times agree fairly well with the observed mean values for temperatures from 20° to 40°C; but beyond 40°C the foregoing analysis appears to be inadequate.

According to Knaysi,[8] "in rod-like bacteria, very active growth results in the formation of long cells which, on account of the practical constancy of the width, are also slender. At this stage, each cell shows several potential places of division, the median one being usually the most advanced. As the velocity of growth diminishes, shorter cells are formed which are uniform in length." If, then, the bacteria are reduced in size by one-half for the most part at 45°C, the coefficient B for them is twice its value for temperatures below 40°C. The generation time, therefore, should be twice the value computed from the low temperature parameters. Thus doubling the value of generation time from the curve at 45.2°C yields 24.2 min, in excellent agreement with the observed value of 24.4 min.

Between 40° and 45°C, during the transition from the long to the short form, the generation time may be expected to have values between one and two times those from the curve. A further halving of V_0 at approximately 47°C would quadruple the value of generation time from the curve, and again fair agreement with observation is obtained. On the other hand, the lengthening of the generation time beyond 40°C, and especially above 45°C, may be entirely due to physicochemical changes that become irreversible when the temperature exceeds 49–50°C.

Other equations with two or more adjustable parameters can be found that fit Barber's data equally well. In fact, Barber showed that a large portion of the curve, representing selected minimum generation times for temperatures between 12.5° and 45°C, approximates a quarter of an ellipse. Kanitz[9] found the reciprocal logarithms of Barber's generation times to be scattered around a straight line with positive slope from 20° to 40°C. This empirical fit approximates Eq. (4), provided it is recalled that the reciprocal of the absolute temperatures is nearly a linear function of the centigrade temperatures between 0° and 50°C. In order to show that our result is more than nice curve fitting, the energy w may be computed from the foregoing value of T^* and applied to multiplication data from a different source. The value of w from Barber's data is equal to 7.45×10^{-13} erg, or nearly half an electron volt.

Multiplication of bacteria.—Suppose a large volume of sterile broth is inoculated with a small number of actively growing bacteria and kept at a uniform temperature for several hours. By a suitable technic the number of bacteria per unit volume may be measured at predetermined intervals of time. Since the numbers become extremely large, the logarithm of the number per unit volume is plotted as a function of time to yield a "multiplication curve."

Often such curves are represented rather well by the logistic function, as first shown for *Escherichia coli* by M'Kendrick and Pai.[10] In one form of this function, the natural logarithm of the ratio of the number of organisms in a unit volume of the medium to the number yet to grow is a linear function of time; that is,

$$\ln[N/(N_\infty - N)] = \mu t + \nu, \qquad (5)$$

where N_∞ represents the number of organisms in a unit volume after a long time ($t\to\infty$), and ν equals $\ln[N_0/(N_\infty - N_0)]$. The coefficient μ is an adjustable parameter. Its dependence on temperature, size of organism, and other factors has

[8] G. Knaysi, *Elements of bacterial cytology* (Comstock, Ithaca, 1944), p. 19.

[9] A. Kanitz, *Temperatur und Lebensvorgange* (Gebrüder Borntraeger, Berlin, 1915).

[10] A. G. M'Kendrick and M. K. Pai, *Proc. Roy. Soc. Edinburgh* **31**, 649 (1911).

not been formulated explicitly hitherto. In the usual applications of the logistic function no insight is gained regarding the details of the multiplication process. In fact, there is some question as to whether the function should be applied at all; for other mathematical relations may be found that can serve equally well.[11–13] In the following discussion the logistic function is shown to be applicable to rod-shaped bacteria that satisfy the foregoing assumptions. Probably it should be considered no more than a first approximation for bacteria of other shapes than those with areas proportional to volumes.

For the multiplication process,

$$N = N_0 2^p \quad \text{and} \quad t = \sum_0^{p-1} \tau_p,$$

where τ_p is the generation time for the pth generation. Here τ_p for all the generations is given by Eq. (3) with ln2 replaced by $\ln[2(N_\infty - N)/(N_\infty - 2N)]$. The summation may be carried out exactly for the case of rod-shaped bacteria. The result is just the logistic function with μ equal to $(n_0 + bN_0)\alpha\bar{c}V_0/2ba$. The quantity $N_0 + n_0/b$ represents the ultimate number of organisms in a unit volume of the medium when all the critical substance is used up just as the final fissions take place. In general, this very special circumstance does not obtain. The number of organisms may be less than half of $N_0 + n_0/b$ when multiplication has nearly ceased.

Lane-Claypon[14] obtained multiplication curves for *Escherichia coli* at a series of temperatures. Generation times were computed from the linear portions of these curves and are entered in the second column of Table I. The third column contains corresponding values of generation times determined from Eq. (4), with T^* set equal to 5560°K and, in this case, B set equal to 1.18×10^{-4} min deg. The agreement exhibited here is encouraging.

Duration of the log phase.—The straight-line portion of the multiplication curve is called the "log phase." Its duration depends on the temperature of the medium, on the amount of inoculum, and on the size of the organism. The upper limit of the log phase usually is chosen to be that point where the multiplication curve diverges noticeably from the straight line portion. The criterion for this point is somewhat arbitrary; a change in slope from that of the straight part should be evident when the number of organisms equals approximately 1 percent of the ultimate number. The time t^* to reach this concentration of bacteria is found by letting N equal $0.01N_\infty$ in Eq. (5); thus, very nearly,

$$t^* = (2ab/\alpha\bar{c}n_0V_0)(\ln N_\infty - \ln N_0 - 4.605). \quad (6)$$

At constant temperature and for uniform conditions of growth, the duration of the log phase is seen to be a decreasing linear function of the logarithm of the amount of inoculum N_0. This relation was found experimentally by Ingraham,[15] who used it to evaluate the bacteriostatic action of dyes. Recently, Hoffmann and Rahn[16] confirmed the relation and noted further that no satisfactory explanation for it had been found. The present analysis suggests that it is a natural result of the multiplication process.

For the log phase of multiplication the generation time is given by Eq. (3). Expressing t^* in hours and τ in minutes, and using common logarithms, one finds the ratio τ/t^* to equal $18.06/\log(0.01N_\infty/N_0)$. Jennison[17] obtained an average value of 5.3 for this ratio directly from multiplication data on *Aerobacter aerogenes*, *Serratia marcescens*, *Chromobacterium violaceum* and two strains of *Escherichia coli*. His cultures were inoculated with 1000–2000 cells per milliliter and had maximum yields near 500 million cells per milliliter. From these values of N_0 and N_∞, the ratio is computed to be somewhat larger than 5. Since this ratio is independent of temperature, the slope of the straight part of the multiplication curve and the duration of the log phase should depend on temperature in a way similar to the generation time.

Finally, it should be noted that, for rod-shaped bacteria, t^* is inversely proportional to the post-fission volume V_0. Because of this dependence, cultures started from old and, therefore, short

[11] R. E. Buchanon and E. I. Fulmer, *Physiology and biochemistry of bacteria* (Williams & Wilkins, 1928–30), vol. 2.
[12] V. A. Kostitzin, *Mathematical biology* (G. G. Harrap, London, 1939), pp. 63–66.
[13] C. G. Lemon, *J. Hygiene* **33**, 495 (1933).
[14] J. E. Lane-Claypon, *J. Hygiene* **9**, 239 (1909).
[15] M. A. Ingraham, *J. Bact.* **26**, 573 (1933).
[16] C. E. Hoffmann and O. Rahn, *J. Bact.* **47**, 177 (1944).
[17] M. W. Jennison, *J. Bact.* **30**, 603 (1935).

cells should be expected to have long log phases. In such cultures, however, the cells pass through an initial transition stage during which they become transformed from the short to the long form. This transition period contributes to the lag phase of the multiplication process.

According to Eq. (6), large amounts of inoculum and elevated temperatures should shorten the lag, as well as the log phase. This effect of the inoculum was first reported by Rahn,[18] and since then has been the subject of studies, such as those of Penfold[19] and Graham-Smith.[20] The effect of temperature, as predicted by Eq. (6), has been observed by Penfold[19] for *Escherichia coli*, by Mueller[21] for *Pseudomonas fluorescens*, and by Stern and Frazier[22] for *Lactobacillus bulgaricus*.

The critical substance.—Up to this point no attempt has been made to identify the critical substance. Many investigators have been puzzled by the fact that multiplication of bacteria practically ceases although there appears to be an adequate supply of nutrient medium and the physical factors for growth are at optimum values. As long as the food content of the medium exceeds a rather low minimum amount, multiplication continues until a maximum yield within a narrow range is reached.[23] This final yield is nearly independent of the amount and age of the inoculum.[22] Although the limitation of the multiplication process in some cultures can be traced to the presence of bacteriocidal chemicals or metabolic products, more often than not multiplication approaches a standstill as the result of an as yet unknown cause. If this cause were the depletion of a critical substance from the culture, that substance probably should be a common one of relatively low molecular weight.

As a matter of fact, the molecular weight of the critical substance can be estimated from the value of B in Eq. (4). On solving for the molecular weight, there results

$$M=(\pi a l_0 k B N_\infty \times 10^{12})^2/5w. \qquad (7)$$

TABLE I. Generation times from multiplication curves for *E. coli*.

Temperature (°C)	Generation times (min) Lane-Claypon	Eq. (4)
20	78	70
25.5	52	48
31	32	34
34.3	27	28
42	19	17

For *Escherichia coli* the quantities in this relation have the values: $w=7.45\times10^{-13}$ erg, $B=1.0\times10^{-4}$ min deg., $a=0.25$ micron, $l_0=1.35$ micron, $N_\infty=1.5\times10^9$ cells/milliliter. The value for N_∞ is the average of 18 determinations by Hershey.[24] The Boltzmann constant k is equal to 1.37×10^{-16} erg deg.$^{-1}$. From these data the molecular weight of the critical substance is found to equal approximately 55. There is considerable experimental evidence that dissolved oxygen is the important limiting factor.[24–29] Although our theoretical result is of the right order of magnitude, we should not conclude that oxygen is the critical substance. Better numerical values of the quantities in Eq. (7) might lead to a molecular weight several times larger, or possibly smaller, than 55.

Many factors not taken into account here may influence bacterial growth and multiplication. For instance, growth promoting or growth inhibiting substances may be produced by bacteria and diffuse into the medium. Such factors have been suggested to account for some of the observations described in the foregoing discussion. If these factors do exist, probably they are of secondary importance. The manner in which the exponential growth of a single cell, Barber's observations on generation time, Ingraham's relation for the duration of the log phase of multiplication, and the effects of the amount of inoculum and temperature on the multiplication process all fit together in the foregoing theory, implies that the kinetic approach to the problem of bacterial growth and multiplication is fundamental.

[18] O. Rahn, *Physiology of bacteria* (Blakiston, 1932), p. 201.
[19] W. J. Penfold, *J. Hygiene* **14**, 215 (1914).
[20] G. S. Graham-Smith, *J. Hyg.* **19**, 133 (1920).
[21] M. Mueller, *Arch. Hyg.* **47**, 127 (1903).
[22] R. M. Stern and W. C. Frazier, *J. Bact.* **42**, 479 (1941).
[23] J. P. Cleary, P. J. Beard and C. E. Clifton, *J. Bact.* **29**, 205 (1935).
[24] A. D. Hershey, *J. Bact.* **37**, 285 (1939).
[25] M. E. Greig and J. C. Hoogerheide, *J. Bact.* **41**, 549 (1941).
[26] C. P. Hegarty, *J. Bact.* **31**, 18 (1936).
[27] A. D. Hershey and J. Bronfenbrenner, *J. Gen. Physiol.* **21**, 721 (1938).
[28] P. P. Levine, *J. Bact.* **31**, 151 (1936).
[29] D. C. Martin, *J. Gen. Physiol.* **15**, 691 (1932).

Introduction to the Elementary Theory of Linear Servomechanisms*

L. Jackson Laslett
Iowa State College, Ames, Iowa

THE theory of control mechanisms of the type known generally as servomechanisms is of interest to workers in many fields. The physicist deals with apparatus in this category when he constructs control equipment or "automatons" for his experimental work, while the engineer applies the same principles when he designs regulator equipment which may involve mechanical, hydraulic, and thermal devices or may be entirely electronic in nature. The applied mathematician can find in this rapidly developing field much that is of interest and that merits study.

The increasing interest in servomechanisms has been the subject of editorial comment,[1] and an introductory article concerning them has recently appeared in the *Journal*.[2] As it is desirable that servomechanism theory be familiar to both engineers and physicists, it seems evident that the elements of such a theory could profitably be presented to college students prior to the time at which some of them undertake advanced study and enter a specialized field of endeavor.

In the conventional analysis of servomechanism behavior frequent use is made of the rather specialized methods of transient analysis. In addition, one finds various stability criterions employed, although their equivalence is scarcely obvious. In order to make the elementary theory of servomechanisms more generally known, and hence more useful, it would appear desirable to attempt to construct the fundamentals of such a theory on as simple a basis as practicable. Such a basis might require a specialized background no more advanced than a general knowledge of differential equations, advanced calculus and, for application of the theory to illustrative problems, elementary analysis of electronic circuits.

In the following sections an attempt will be made to give a simple treatment of linear servomechanisms, with particular attention to the stability problem. No description will be given of the multitude of control devices to which the theory might be applied, but it is hoped that this treatment will serve to integrate in a simple way some of the different types of approach that one encounters in this field and so will be of value to those becoming interested in the subject.

1. Definition of a Servomechanism[3]

In a straightforward *control mechanism*, an input x_1 is deliberately given a value or series of values that cause the mechanism to produce an output x_2, which in general will vary with time but which it is intended shall show, at least approximately, some desired relationship to the input. The quantities x_1 and x_2 can be of a wide variety of types, such as mechanical displacement (linear or angular), electric potential (or current), or temperature, to mention only a few. It is not necessary, of course, that x_1 and x_2 be of the same nature. It will be realized that the control mechanism will in general be provided with a source of power, so that the control signal may be quite weak and yet control a unit requiring considerable power. Thus the valve on a gas range, which requires little energy to adjust and none to hold in position, is a control for the temperature above the burner and provides an illustration of a control mechanism even more homely than a familiar electronic amplifier (without feedback), which also falls in this category.

In a *general servomechanism* an important addition is made to a control mechanism of the "open-cycle" type just described. The mechanism is so arranged that the input is modified or corrected by the output. Thus, customarily, the input is compared with the output, or with some quantity generated by the output, and the re-

* Paper No. 23 from the Institute for Atomic Research, Iowa State College.

[1] *Electronics* **20**, 74 (May, 1947).

[2] R. H. Bacon, *Am. J. Physics* **16**, 79 (1948). After preparation of the manuscript for the present article, the author learned that Bacon's paper was to appear. The present article has accordingly been somewhat condensed from its original form. The speed controller, for which an analysis is given in Sec. 3, may be compared with the second example cited by Bacon.

[3] A discussion of the nature of a servomechanism has been given by Bacon, reference 2. The statement given here will serve as an introduction to the approach used in the remainder of the article.

sultant difference signal is employed as the input to the control mechanism. The distinction between a simple control mechanism and a more general servomechanism is indicated by the block diagrams of Fig. 1. It is thus seen that the provision of a feedback path (or, more generally, of any number of feedback paths) results in an effective input that depends upon what the mechanism is doing. The device is what is known as a general servomechanism, of which a straight control mechanism without feedback might be considered a degenerate case.[4]

Although it is clear that the provision of a feedback path provides an automatic means of monitoring and perhaps improving the performance of a control mechanism, an evaluation of the merits of such an arrangement requires a quantitative examination of the effect of feedback upon performance. There are a number of points to consider: (i) whether the system will be unstable and will oscillate; (ii) with what faithfulness the output follows the input, which may be constant or some function of time; and (iii) how sensitive the accuracy of following is to those parameters, such as power-supply voltage and the like, that affect the characteristics of the mechanism.

2. Elementary Analysis[5]

(*A*) *Nomenclature and basic relationships.*—We will designate the input to the servomechanism

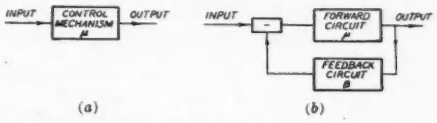

FIG. 1. Block diagrams of: (*a*) a simple control mechanism; (*b*) a closed-cycle servomechanism.

by the function $x_1(t)$ and the output by $x_2(t)$, where t represents the time. The output function depends, of course, on the nature of the input function in a manner dependent upon the various parameters of the system. In the absence of any feedback loop we can regard $x_2(t)$ as the result of the application to $x_1(t)$ of an operator $\mathfrak{y}$, the transfer function, which characterizes the direct control system [Fig. 1(*a*)]. If we restrict ourselves to systems which, by a suitable selection of the quantities x_1 and x_2, follow *linear* differential equations, the relationship between these quantities may be written formally as

$$x_2(t) = [\mathfrak{y}(\mathbf{p})]x_1(t), \tag{1}$$

for no feedback. Here $\mathbf{p}$ is the symbol taken to represent differentiation with respect to time[6] and, for the cases considered here, $\mathfrak{y}$ may be regarded as a fraction containing in the numerator and denominator polynominals in $\mathbf{p}$.

As a first illustration of a relationship of the form of Eq. (1), consider the equation governing the temperature of a calorimeter of thermal capacity C, to which energy is being supplied at a given rate. We shall let x_1 represent the power supplied and x_2 the temperature of the calorimeter with respect to its surroundings. If we now assume the rate of heat loss to these surroundings to be given, in accordance with Newton's law of cooling, by Kx_2, the following differential equation will apply:

$$C\mathrm{d}x_2/\mathrm{d}t = x_1 - Kx_2.$$

If we write this relation in the form

$$x_2 = x_1/(K + C\mathbf{p}),$$

a comparison with the general relationship (1) gives us,

[4] Although we have here emphasized the distinction between a true servomechanism of the general type and the more rudimentary control mechanism without feedback, in a certain sense most simple control equipment could be regarded as possessing a feedback in that usually there is an operator who can modify the input at his discretion, as suggested by any observed departure from the desired performance. It will be readily appreciated, however, that such a human link in many cases would not only represent a wasteful way in which to employ a man's time, but that in addition this link may be incapable of giving the rapid, accurate and reliable performance that may be required, and will exhibit troublesome nonlinearities.

[5] It is appropriate to acknowledge at this point the indebtedness of the present author to the excellent treatment of this subject by L. A. MacColl, *Fundamental theory of servomechanisms* (Van Nostrand, 1945) and to mention in addition the thorough work of H. W. Bode, *Network analysis and feedback amplifier design* (Van Nostrand, 1945). While the manuscript for this article was in preparation two additional books on the subject appeared: H. Lauer, R. Lesnick and L. E. Matson, *Servomechanism fundamentals* (McGraw-Hill, 1947), and H. M. James, N. B. Nichols and R. S. Phillips (Editors), *Theory of servomechanisms* (McGraw-Hill, 1947).

The following general references should also be cited: N. Minorsky, *J. Frank. Inst.* **232**, 451, 519 (1941); R. E. Graham, *Bell Syst. Tech. J.* **25**, 616 (1946); A. C. Hall, *J. Frank. Inst.* **242**, 279 (1946); A. C. Hall, *Analysis and synthesis of linear servomechanisms* (Massachusetts Institute of Technology Press, 1943); E. S. Smith, *Automatic control engineering* (McGraw-Hill, 1944).

[6] This notation is introduced here primari-ly for convenience. The extent to which this operator can be treated by the usual rules of algebra is shown in such books on differential equations as the following, in which the symbol D is employed for our operator $\mathbf{p}$: A. R. Forsyth, *A treatise on differential equations* (Macmillan, ed. 4, 1914), chap. III; H. T. H. Piaggio, *Differential equations and their applications* (Bell, 1929), chap. III; L. M. Kells, *Elementary differential equations* (McGraw-Hill, ed. 2, 1935), chap. VI; F. H. Miller, *Partial differential equations* (Wiley, 1941), chap. I.

for this example,

$$\mathfrak{y}(\mathbf{p})=1/(K+C\mathbf{p}).$$

A second, simple illustration similar to the preceding is that of an idealized direct-current shunt generator, for which we require the relation between the open-circuit output voltage e_2 and the voltage e_1 impressed across the field windings. We assume that the e.m.f. generated is equal to a constant k times the field current i_f, and relate this field current to the input voltage by the equation

$$L(\mathrm{d}i_f/\mathrm{d}t)+Ri_f=e_1,$$

where L and R are, respectively, the inductance and resistance of the generator field. The relation connecting e_2 with e_1 then becomes

$$e_2(t)=\frac{k}{R+L\mathbf{p}}e_1(t),$$

so that

$$\mathfrak{y}(\mathbf{p})=k/(R+L\mathbf{p}).$$

It will be noted that the expression for $\mathfrak{y}(\mathbf{p})$ in this case and in the previous example involves a factor of the form $1/(1+T\mathbf{p})$, the single time constant T being given by C/K in the first example and by L/R in the second.[7]

When one introduces a feedback loop (or, as may occur in practice, several such loops in parallel) into the system, the input is supplemented with a signal derived from the output [Fig. 1(*b*)]. The output $x_2(t)$ is acted upon by a second transfer function $\mathfrak{z}(\mathbf{p})$, and the result is combined with the input signal in such a manner that the effective input now becomes $x_1(t)-[\mathfrak{z}(\mathbf{p})]x_2(t)$. Accordingly Eq. (1), which applied in the absence of feedback, must now be replaced by the relation

$$x_2(t)=[\mathfrak{y}(\mathbf{p})][x_1(t)-[\mathfrak{z}(\mathbf{p})]x_2(t)], \qquad (2)$$

or more concisely,

$$x_2(t)=\frac{\mathfrak{y}(\mathbf{p})}{1+\mathfrak{y}(\mathbf{p})\mathfrak{z}(\mathbf{p})}x_1(t). \qquad (3)$$

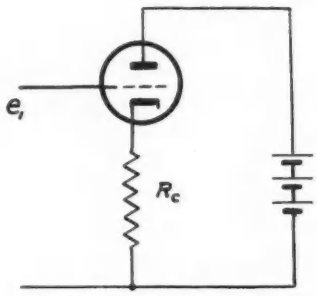

FIG. 2. Cathode-follower circuit.

As an illustration of the application of the notation just introduced, consider the case of the cathode-follower circuit, illustrated in Fig. 2, to which the term "current-feedback" is sometimes applied. To introduce a bit of variety, let us consider as the output in which we shall be interested the *change* (from its value when $e_1=0$) of the *current* in R_c and the tube. If the signal e_1 were applied directly between the grid and cathode, we would have a simple, or open-cycle, control mechanism, for which linear analysis gives $i=\mu e_1/(R_p+R_c)$. Here R_p represents the plate resistance of the tube, and μ is the amplification factor—not to be confused with the general operator $\mathfrak{y}(\mathbf{p})$, which we have previously introduced to represent an overall transfer function.

If, in the foregoing example, we identify e_1 and i, respectively, with x_1 and x_2 of our general discussion, we have

$$\mathfrak{y}(\mathbf{p})=\mu/(R_p+R_c).$$

Since actually, however, we have a grid-cathode signal that is not e_1, but e_1-iR_c, we have a feedback loop for which, in this notation, $\mathfrak{z}(\mathbf{p})=R_c$. Our analysis then leads to the well-known result

$$i(t)=\frac{\mathfrak{y}(\mathbf{p})}{1+\mathfrak{y}(\mathbf{p})\mathfrak{z}(\mathbf{p})}e_1(t)$$

$$=\frac{\mu/(R_p+R_c)}{1+\mu R_c/(R_p+R_c)}e_1(t)=\frac{e_1(t)}{R_c+(R_p+R_c)/\mu}.$$

For this particular example it will be observed that the operators $\mathfrak{y}(\mathbf{p})$ and $\mathfrak{z}(\mathbf{p})$ are mere constant multipliers and do not actually involve $\mathrm{d}/\mathrm{d}t$.

Attention might well be directed at this point to some tacit assumptions made in the foregoing general treatment. One such assumption is that the *loading* introduced into the input and output circuits by the addition of the feedback circuit is insufficient *per se* to affect $\mathfrak{y}$ significantly; if the loading does affect $\mathfrak{y}$ a new expression for this transfer function must be used. A second assumption is that the buffering is sufficiently complete in the feedback circuit that, with the gain which is present in the direct circuit, the input cannot influence the output appreciably by the

[7] It should be mentioned that the relation obtained here for the output of a shunt generator applies only to the idealized case in which the value of T is unaffected by such phenomena as eddy-current loss and nonlinearities in the iron, but the derivations given should serve to illustrate the notation introduced to this point. From the foregoing examples it will be appreciated that the quantities x_1 and x_2 need not necessarily be similar dimensionally and that therefore the operator $\mathfrak{y}$, which represents the direct transfer function of the system, will not in general be dimensionless but may be regarded as a sort of generalized admittance. It nevertheless appears desirable to retain the use of the symbol $\mathfrak{y}$ for this operator, since it is suggestive of a generalized gain or amplification factor and is frequently so used in this theory when specifically applied to electronic circuits.

transmission of a signal through the feedback circuit in a direction opposite to that which the feedback signals follow. It should also be remarked that some writers define the operator $\mathfrak{B}$ as the *negative* of the one which we have so designated throughout the present article.

(*B*) *Investigation of stability.*—(1) Before one proceeds with an investigation of such questions as the accuracy with which a servomechanism will follow a given input signal, it is prudent to inquire whether, by virtue of the feedback path that has been provided, the device can be unstable. It will be recognized, of course, that it is quite possible for such a system to be unstable, and it becomes important to determine under what conditions this situation will occur. Although this question may be attacked in a variety of ways, it may be regarded as a question of the behavior of the solutions of the applicable differential equation. We will here confine our attention to those mechanisms for which the feedback may be represented by a single feedback path, although, as MacColl[5] has indicated, quite similar results may sometimes be applicable to cases of multiple-feedback servomechanisms.

If we rewrite Eq. (2) in the form

$$[1+\mathfrak{y}(\mathfrak{p})\mathfrak{B}(\mathfrak{p})]x_2(t)=[\mathfrak{y}(\mathfrak{p})]x_1(t), \qquad (4)$$

we have the differential equation for $x_2(t)$. Whether the input $x_1(t)$ be constant or exhibit some particular variation with time, we have, as is well known, in addition to some particular solution, the complementary function involving the sum of terms $C_m e^{p_m t}$, where the C_m are constants and the p_m are roots of the algebraic equation

$$[1+\mathfrak{y}(p)\mathfrak{B}(p)]=0, \qquad (5)$$

and will in general be complex. For such terms in the complementary function as arise from a multiple root of Eq. (5), a factor which is a polynominal in t must be appended to the exponential. To avoid having, for any set of initial conditions, terms in the solution which grow indefinitely with time, *one requires that no roots of Eq. (5) lie in the positive half of the complex plane.*

Roots lying in the negative half of the plane will introduce transient terms only into the solution of Eq. (4), although for the output to follow the input reasonably well, it is desirable that these transient terms die out rapidly and therefore that the magnitude of the real part of each p_m be reasonably large. A root at the origin may not be objectionable in some applications. If, however, the parameters of the system happened to be such that a complex root of Eq. (5)—and its complex conjugate, which would also be a root—should lie *exactly* on the axis of imaginaries, we would have the possibility of a sustained oscillation. This not only could be undesirable in itself, but might lead to oscillations of increasing amplitude if the system parameters should change even slightly. A *multiple* root on this axis would, of course, make possible oscillations which increase with time and so would imply instability.

By the application of tests such as Routh's rule,[8] it is possible to determine whether or not there are roots of Eq. (5) in the positive half of the complex plane, without making a specific determination of the value of each root. It is evident that Routh's rule can be very helpful in examining the stability of a given system, but it is believed that an alternative method of analysis, which will be discussed next, will often be more instructive as a guide to system design.

(2) There is a second formulation of the criterion for stability, which perhaps can be made to appear plausible directly, but which can also be deduced from the criterion given above. This alternative formulation eliminates the necessity of studying directly the location of the roots of Eq. (5) for each set of system parameters that one may wish to consider and perhaps gives one a better presentation of the properties of the system. The method makes use of the so-called Nyquist graph,[9] for which one considers the ex-

[8] E. J. Routh, *A treatise on the dynamics of a system of rigid bodies* (Macmillan, ed. 6, 1905), vol. 2, chap. VI. Compare also the less used but concise formulation of an equivalent criterion given by A. Hurwitz, *Math. Ann.* **46**, 273 (1895).

The use of Routh's rule is described by M. F. Gardner and J. L. Barnes, *Transients in linear systems* (Wiley, 1942), vol. 1, chap VII, sec. 9. Although this latter work is primarily devoted to the development of specialized methods of operational analysis, some of the problems given could serve as useful examples for the illustration of the principles discussed in this paper.

[9] H. Nyquist, *Bell Syst. Tech. J.* **11**, 126 (1932). See also E. Peterson, J. G. Kreer and L. A. Ware, *Proc. IRE* **22**, 1191 (1934), in which the Nyquist criterion applied to an amplifier is shown to be expressible in terms of the behavior of a polar graph of the input impedance.

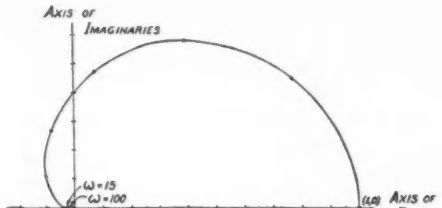

FIG. 3. Nyquist graph of $\mu\beta/A$ for the simple speed regulator of Fig. 6(*a*).

pression $1+\mu(p)\beta(p)$ as a function of ω, where $p \equiv j\omega$, and plots in the complex plane the values that this function assumes for all real values of the variable ω. This corresponds to evaluating the function for points along the axis of imaginaries in the complex p-plane and will be understood to include distorting infinitesimally the path in the p-plane so as to make it pass to the right of the origin, if this be necessary to avoid ambiguities due to a possible singularity at that point. It is also to be understood that the path at infinity is completed by an "infinite semicircle" lying in the positive portion of the p-plane. This latter specification will not be needed, of course, for those cases in which the function approaches, for large values of the independent variable, a nonzero value that is independent of the manner in which the limit is approached.

What is being considered here is, then, evidently a transformation of a specific closed curve in the p-plane into a closed curve in the $[1+\mu(p)\beta(p)]$-plane. If we were first to start with some arbitrary point in the p-plane—a point that is neither a root nor a pole of $1+\mu(p)\beta(p)$—and pass around a sufficiently small closed contour in that region of the plane, a corresponding small closed curve would be traced out through the corresponding point in the $[1+\mu(p)\beta(p)]$-plane and would *not* encircle the origin. If the path in the p-plane were expanded and modified, that in the other plane would also be affected but the situation might appear essentially unchanged in a qualitative way. If, and only if, we now come to encircle in the p-plane a root p_m of Eq. (5), will the curve in the $[1+\mu(p)\beta(p)]$-plane come to encircle the origin. This can be seen by recognizing that the presence of a factor $(p-p_m)$ in the expression $1+\mu(p)\beta(p)$ will then result in the argument (angle) of this latter expression changing by 2π in going once around the contour. In this discussion we have left out of consideration the effect of encircling in the p-plane points that are poles of the expression $1+\mu(p)\beta(p)$; if such poles were present in the portion of the complex plane under consideration, the effect of encircling them would be, of course, to cause the transformed curve to encircle the origin in the reversed direction. It will be recognized that the considerations outlined here do no more than permit us to avoid referring explicitly to a theorem due to Cauchy,[10] which gives us more directly the result just stated.

It can now be seen that the condition previously given for stability—namely, that there be no roots of Eq. (5) in the positive half of the p-plane (that is, within our specific contour)—is equivalent to the condition that the *Nyquist graph should not encircle the origin*. If we plot merely $\mu\beta$ (instead of $1+\mu\beta$) as a function of ω, we are led to the alternative formulation that *the* $\mu\beta$ *graph should not encircle the point* -1. This is Nyquist's criterion for stability.

Frequently the result of various changes in system parameters can be seen readily from a $\mu\beta$ graph. In many cases, for example, the graph will be found to be of such a nature that, as plotted, it does not encircle the Nyquist point -1, and so implies stability, but if the over-all gain be increased (increasing μ by a constant, real factor), this critical point will be encircled and instability will occur. A graph of this nature is exhibited in Fig. 3 in connection with an example treated in SEC. 3. If for such cases the permissible gain is not adequate for the results desired, the curve must be modified, by changes in design, in such a way as to eliminate or ameliorate this difficulty. Rather exceptional cases have also been described[11] in which a reduction of gain can result in the onset of an unstable condition; the stability which then prevails at high gain

[10] See E. G. Phillips, *Functions of a complex variable* (Oliver and Boyd, ed. 4, 1946), chap. IV, sec. 40; E. T. Whittaker and G. N. Watson, *A course of modern analysis* (Cambridge, 1927), secs. 6.3–6.31; C. Jordan, *Cours d'analyse* (Gauthier-Villars, 1913), vol. 2, sec. 329.

[11] See, for example, the second paper cited in reference 9; also H. Nyquist, U. S. Patent No. 1,915,440.

is a conditional stability and, as can easily be seen, might be objectionable from a practical standpoint.

If we exclude from consideration cases of conditional stability, it is interesting to observe that Nyquist's criterion can be made to seem quite plausible if we first consider the physical meaning of the operator $\mu(\mathbf{p})\beta(\mathbf{p})$. One notes that, save for a minus sign, this operator is the loop gain of the servomechanism. By this is implied that if we break the circuit at the input to the control mechanism and apply at this point a signal represented by $A_i e^{j\omega t}$, the signal returning from the output via the feedback path is $A_i[(-)\mu(j\omega)\beta(j\omega)]e^{j\omega t}$, where $j\omega$ has, of course, been inserted in place of $\mathbf{p}$. If, then, for some real value of ω this return signal should be equal to the original $A_i e^{j\omega t}$ and the circuit loop had been unbroken, conditions would be such that an oscillation of angular frequency ω could be sustained without the application of an independent input to the servomechanism as a whole.

The case just described would occur if $\mu(j\omega)\beta(j\omega)$ should be equal to -1 and corresponds, of course, to the $\mu\beta$-graph passing exactly through the point -1. If the loop gain were a real number greater than unity for some particular value of ω, it might be expected that the system would go into an oscillation of increasing amplitude until limited by nonlinearities in its structure that were not included in our analysis. To avoid excessive transient oscillations, as well as possible undamped oscillations resulting from slight departures of the system parameters from the values used in the analysis, it is desirable that the graph of loop gain shall not even pass *close* to the critical point. For simple cases the foregoing considerations suggest, therefore, the requirement that at frequencies for which the phase shift around the loop is π radians (for $\mu\beta$, as defined here), the magnitude of the loop gain shall be less than unity by a reasonable amount and that when the magnitude becomes greater than unity the phase shift shall be somewhat less than π radians.

(3) From the foregoing type of reasoning it would not be surprising that separate graphs of the magnitude and phase of the loop gain as a function of frequency could be helpful. The relationship between these two curves, which are by no means independent, has been discussed thoroughly by Bode,[12] and a careful treatment will not be attempted here. It would be suitable, however, to observe that in the typical case the loop gain will contain factors of the form $1/(1+j\omega T_m)$, where the T_m are positive time constants characterizing the various inertias (electrical, mechanical, or thermal) of the system. In passing from a value of ω small compared to $1/T_m$ to a value large compared to this quantity, such a factor will introduce a phase change of $-\pi/2$. Similarly a factor $1+j\omega T_m'$ can introduce a phase change of $+\pi/2$, and a factor of the form $(1+j\omega T_m')/(1+j\omega T_m)$ can contribute a bump approaching plus or minus $\pi/2$ in the phase graph, depending upon whether T_m' is greater or less than T_m. As MacColl[13] has suggested, such considerations can be helpful in determining the form of the curves in question. It can further be observed that if gain is plotted as a function of frequency, each factor of the type $1/(1+j\omega T_m)$ introduces, as its maximum contribution to the *slope* when a double logarithmic graph is used, a factor $\frac{1}{2}$ for every factor 2 in frequency. This slope, which in electrical applications is frequently expressed as "-6 db/octave,"[14] would be associated with the phase change of $-\pi/2$ mentioned above.

A direct connection, suggested by the remarks of the preceding paragraph, does not, in fact, hold between the phase shift at some particular frequency and the slope of the gain or attenuation curve at that point. The phase at a given frequency will actually be affected somewhat by the *entire course* of the attenuation curve. An examination of the exact relationship[12] which holds for minimum-phase-shift networks will show, how-

[12] H. W. Bode, *Bell. Syst. Tech. J.* 19, 421 (1940) and U. S. Patent No. 2,123,178. For a short, informative résumé of these principles as they apply to amplifier design, see F. E. Terman, *Radio engineers' handbook* (McGraw-Hill, 1943), sec. 3, par. 27 and sec. 5, par. 11, or his earlier *Radio engineering* (McGraw-Hill, 1937), secs. 51–52. A helpful discussion of the application of feedback to amplifier design is given in *The radiotron designer's handbook* (Wireless Press, 1940), chap. 6, which has been distributed in the U. S. A. by the R. C. A. Manufacturing Co., Harrison, N. J. Note also the earlier paper by Y. W. Lee, *J. Math. and Physics* 11, 83 (1932), especially Eq. (14) of that paper, relating the phase shift explicitly to the amplitude characteristics of an admittance or transfer function.

[13] MacColl, reference 5, chap. VII.

[14] It should perhaps be mentioned that the decibel (db) is a logarithmic unit properly applied to power. We here take the number of decibels as given by 20 times the common logarithm of the transmission factor.

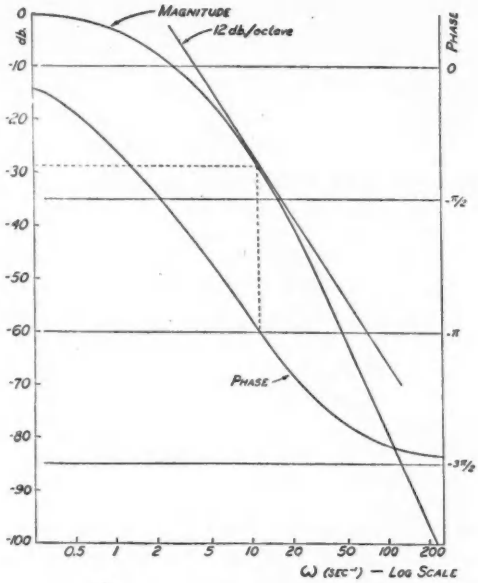

FIG. 4. Graphs of the magnitude and phase of $\mu\beta/A$ for the simple speed regulator of Fig. 6(*a*).

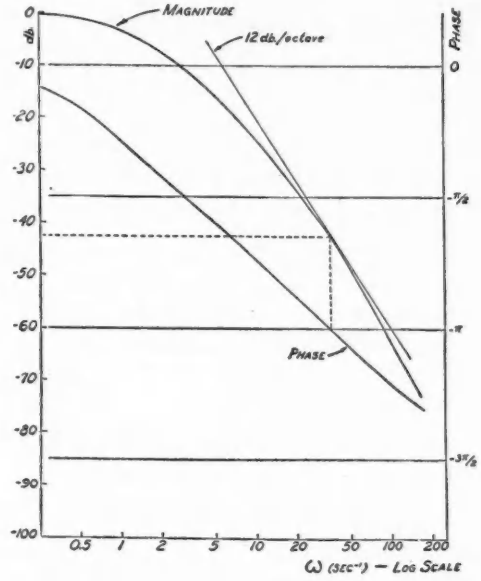

FIG. 5. Graphs of the magnitude and phase of $\mu\beta/A$ for the system of Fig. 6(*a*) with a filter of the type shown in Fig. 6(*b*) added.

ever, that the phase shift at a particular frequency depends *primarily* upon the slope of the transmission curve at that point and may be taken as approximately $\pi/12$ *rad*/(*db*/*octave*). We have already seen that, for those frequencies for which the phase shifts total $-\pi$, we would normally require that the loop gain be less than unity (that is, below 0 db). From the relation just mentioned it then becomes natural to require, on the basis of a loop-gain graph alone, that *a rate of fall of 12 db/octave or more be avoided for gains above the zero-decibel level.*[15]

It is on the basis of considerations such as these that one comes to realize that the provision of a wide band in a high-gain feedback amplifier requires that the response be controlled over a very large *additional* range of frequencies in order that oscillations be avoided. In such cases the consideration of the effect of parasitic circuit elements frequently becomes essential. As a means for improving the transmission characteristics of a servo system, a design modification that will introduce a factor $(1+j\omega T_m')/(1+j\omega T_m)$ into the expression for the loop gain can be helpful. The bumps in the phase graph which will be contributed by a factor of this form may result in the avoidance of excessive negative phase shifts in the frequency region for which the loop gain is high. Gain and phase curves, with a specific illustration of the effect of a factor of this type, are given in Figs. 4 and 5 in connection with an example to be discussed later.

(*C*) *Faithfulness of following.*—We come now to consider how well a servomechanism (assumed stable) gives the output desired. Frequently what is wished is merely that $x_2(t)$ *reproduce*, at a higher power level, the input $x_1(t)$, but we may with greater generality take the desired relationship as

$$x_2^0(t) = [\mathbf{f}(\mathbf{p})]x_1(t), \tag{6}$$

where $x_2^0(t)$ represents the *ideal* output corresponding to the input $x_1(t)$, and $\mathbf{f}(\mathbf{p})$ symbolizes the desired relationship between these two quantities.

If Eq. (6) were satisfied *exactly* by $x_2(t)$ itself, reference to Eqs. (2) or (3) shows that we would then have, *ideally*,

$$[1+\mu(\mathbf{p})\beta(\mathbf{p})][\mathbf{f}(\mathbf{p})] = \mu(\mathbf{p}). \tag{7}$$

[15] See Bode, reference 12, in which a cut-off rate of 10 db/octave is suggested as a frequently suitable upper limit to this slope for gain values above 0 db.

It is customary to attempt to satisfy this condition closely by seeking a design such that, *for those frequencies for which it is required that the response be in accord with relationship* (6),

$$\mu(j\omega)\beta(j\omega) \gg 1 \tag{8a}$$

and

$$\beta(j\omega)\mathbf{f}(j\omega) = 1. \tag{8b}$$

It will be observed that the equality (8b) thus required does not include μ. This is of advantage since the μ-circuit may involve power elements whose response may be somewhat intractable to exact adjustment, whereas the β-circuit can be composed of units operating at a low power level and capable of adjustment to a condition which they can then be relied upon to maintain with precision.

The *relative error* in the performance of the servomechanism may be written

$$\text{rel. error} \equiv [x_2(t) - x_2^0(t)]/x_2^0(t) = \frac{x_2(t) - [\mathbf{f}(\mathbf{p})]x_1(t)}{[\mathbf{f}(\mathbf{p})]x_1(t)}. \tag{9}$$

If condition (8b) is satisfied, this becomes, through the use of Eq. (3),

$$\text{rel. error} = \frac{-1}{1+\mu(j\omega)\beta(j\omega)} \cong \frac{-1}{\mu(j\omega)\beta(j\omega)}, \tag{10}$$

for simple-harmonic inputs. It is then seen that, if condition (8a) is also satisfied, this relative error will be small. Its amount depends upon the loop gain and so upon the amplification in the direct μ-circuit, as well as upon the feedback β. The variation of output arising from variations in parameters that affect μ will be discussed quantitatively in SEC. 2 (*D*). Improved stability and a reduction of distortion are two important advantages that can be gained by the use of feedback in amplifier design.[16]

It is also of interest to consider cases in which the desired output has some simple form other than simple harmonic and to investigate the departure of the actual output from the ideal. The nature of the output can, of course, be sought by a study of the original differential equation, in which, by Eq. (6), $x_1(t)$ would be expressed in terms of $x_2^0(t)$. In certain cases the solution that is valid for large values of t, after the transient terms have died out, can be expressed quite simply in terms of the coefficients of the following formal series representation for the operator, which relates the actual output $x_2(t)$ to the desired output $x_2^0(t)$:

$$\frac{\mu(\mathbf{p})/\mathbf{f}(\mathbf{p})}{1+\mu(\mathbf{p})\beta(\mathbf{p})} = \mathbf{p}^{-n}[a_0 + a_1\mathbf{p} + a_2\mathbf{p}^2 + \cdots]. \tag{11}$$

The results of an analysis of this character for a few simple functions will be summarized here. For an input such that the desired output $x_2^0(t)$ is a constant A_0, a satisfactory steady-state solution of the differential equation is obtained if $n=0$ and $a_0 \cong 1$. This solution for $x_2(t)$ is, namely, a_0A_0. If, as a second example, $x_2^0(t)$ is to be a "ramp function,"[17] which assumes the form A_1t for $t \geq 0$, the limiting solution for $x_2(t)$ will be (with $n=0$) $a_0A_1(t+a_1/a_0)$. Clearly it is desirable in this case that not only $a_0 \cong 1$, but that a_1 be small numerically. The term a_1A_1 in this solution could be regarded as a velocity error and a_1 termed the velocity-error coefficient. Lastly, for a desired output of the form A_2t^2 it is found that not only the magnitude of a_1 but also that of $2a_2 - a_1^2$ should be made small.

It will be observed that in the case of the cathode-follower circuit considered for illustrative purposes in SEC. 2 (*A*), the value of a_0 may be written

$$a_0 = 1 - \frac{R_p + R_c}{\mu R_c + (R_p + R_c)}$$

and so will be close to unity for any large value of μ, whereas in this elementary analysis the remaining coefficients $a_1, a_2, \cdots$ vanish. Actually, of course, one might be satisfied in this instance with a mere *proportionality* between e_1 and e_2, which requirement would be rigorously met, according to our earlier analysis, so long as the characteristics of the μ-circuit, as well as those of the β-circuit, remain unchanged.

(*D*) *Influence of variations of circuit parameters.*—If the condition (8b) is satisfied, so that for a simple-harmonic input of frequency ω the relative

[16] For a more extensive discussion of these advantages see, for example, H. S. Black, *Bell Syst. Tech. J.* **13**, 1 (1934) or *Elec. Eng.* **53**, 114 (1934), and other references listed in sec. 5, par. 11 of Terman's *Handbook* (reference 12).

[17] This vivid name for the function with which we are here concerned has been ascribed by MacColl (reference 5) to K. K. Darrow. The type of analysis outlined here is described in greater detail by MacColl in chapter VI, reference 5.

error is given by Eq. (10), the effect of a variation in the transfer function of the μ-circuit is readily seen. This can be conveniently written in the following manner:

$$\text{rel. change in output/rel. change in } \mu = \text{d (rel. error)}/(\text{d}\mu/\mu) = \mu(j\omega)\beta(j\omega)/[1+\mu(j\omega)\beta(j\omega)]^2 \quad (12a)$$

$$\cong 1/\mu(j\omega)\beta(j\omega). \quad (12a')$$

A similar result is obtained (i) if, instead of establishing the condition represented by Eq. (8b), we adjust β so that the error vanishes for the value of μ obtaining at the moment and then consider the effect of a variation of μ, or (ii) if we measure the change in output in terms of the *original* rather than the *ideal* output value. Under either of these circumstances one will obtain the following result, which approximates to $1/\mu\beta$ as before:

$$\text{rel. change in output/rel. change in } \mu = 1/[1+\mu(j\omega)\beta(j\omega)] \quad (12b)$$

$$\cong 1/[\mu(j\omega)\beta(j\omega)]. \quad (12b')$$

If the change in μ is due to a change in some parameter ξ of the system (such as a power-supply voltage), the relative change in output per relative change in the parameter can be obtained, of course, by multiplying this expression by the logarithmic derivative $(\text{d}\mu/\mu)/(\text{d}\xi/\xi)$.

3. Illustration of Foregoing Types of Analysis

It will be illustrative to analyze a somewhat idealized servomechanism by the procedures aforedescribed. For this purpose we consider a simplified speed-control circuit for a small permanent-field direct-current motor, whose armature current is provided by an amplifier of gain G. There will be a feedback voltage from a tachometer generator driven by the motor shaft. The e.m.f. developed by this latter machine will be assumed to be *accurately proportional* to the shaft speed, and a fraction of this signal will be combined with the input to the servomechanism to provide the input to the main amplifier. A diagram of this mechanism is shown in Fig. 6(a). It should perhaps be noted that the method indicated for combining the input and feedback signals is not necessarily the most desirable arrangement, since, unless the input impedance of the amplifier is very large, the input signal to the amplifier will not be just the difference of the input and feedback e.m.f.'s but will be influenced by the other resistances and reactances in the circuit. Rather than attempt to take account of such effects in our formulas for $\mu(\mathbf{p})$ and $\beta(\mathbf{p})$, however, we shall in what follows assume that the desired combination of signals has been achieved by a suitably refined arrangement.

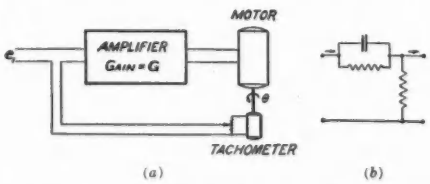

FIG. 6. Schematic diagrams of: (a) a simplified speed-control mechanism; (b) a possible filter whose application is discussed in the text.

In proceeding with the analysis, we shall assume no significant time constants to be introduced by the amplifier itself, but shall consider here only those delays introduced by the motor and, conceivably, by the tachometer output circuit. We might conveniently identify the output function x_2 with the velocity $\text{d}\theta/\text{d}t$ of the motor shaft, but we shall choose instead, for purposes of illustration, to identify x_2 with the angle θ through which the shaft has turned. For x_1 we shall take the input voltage e_1 to the mechanism.

To get the relation between the motor response and the amplifier input, we first assume that the armature current i_a may be related to the amplifier input, in the absence of feedback, by the relation

$$(L\mathbf{p}+R)i_a = Ge_1 - k\mathbf{p}\theta,$$

where L and R are the inductance and resistance of the motor armature circuit as fed by the amplifier, G is the gain of the amplifier, and k is a back-e.m.f. coefficient relating the induced e.m.f. to the armature speed. Since the motor torque is given by the product of the armature current and a factor which we shall designate by k', we also have

$$k'i_a = (J\mathbf{p}^2+\zeta\mathbf{p})\theta,$$

where J and ζ are, respectively, the inertia and coefficient of dynamic friction for the armature and its load. Elimination of i_a from these equa-

tions then gives

$$(1+a\mathbf{p}+b\mathbf{p}^2)\mathbf{p}\theta=(G/K)e_1,$$

where

$$a\equiv(L\zeta+RJ)/(kk'+R\zeta),\quad b\equiv LJ/(kk'+R\zeta),$$

and

$$K\equiv k+R\zeta/k'.$$

The transfer function $\mathfrak{y}(\mathbf{p})$ therefore assumes the form

$$\mathfrak{y}(\mathbf{p})=(G/K)/\mathbf{p}(1+a\mathbf{p}+b\mathbf{p}^2).$$

The feedback circuit can be represented by means of the operator

$$\mathfrak{z}(\mathbf{p})=c\mathbf{p}/(1+T_\beta\mathbf{p}),$$

where the parameter T_β is introduced to take account of any time delay that may unavoidably be present in the tachometer response as received at the amplifier.

If it is desired that the shaft velocity be equal to the product of a specified constant s and the input voltage, the operator $\mathbf{f}(\mathbf{p})$ of Eq. (6) assumes the form

$$\mathbf{f}(\mathbf{p})=s/\mathbf{p}.$$

If we then proceed with the design in a manner consistent with Eqs. (8) for $p=0$, we adjust the magnitude of the feedback signal so that $c=1/s$. We then have

$$\mathfrak{y}(\mathbf{p})\mathfrak{z}(\mathbf{p})=\frac{A}{(1+a\mathbf{p}+b\mathbf{p}^2)(1+T_\beta\mathbf{p})},$$

where we have written $A\equiv G/Ks=\mathfrak{y}(0)\mathfrak{z}(0)$.

To permit an illustrative numerical calculation, let us assign the following arbitrary, and not necessarily realistic, values: $a=8/7$ sec, $b/a=\frac{1}{8}$ sec, and $T_\beta=\frac{1}{16}$ sec. Then

$$\mathfrak{y}(\mathbf{p})\mathfrak{z}(\mathbf{p})=\frac{A}{(1+\mathbf{p})(1+\mathbf{p}/7)(1+\mathbf{p}/16)}.$$

A Nyquist polar graph of the function $(1/A)\mathfrak{y}(j\omega)\mathfrak{z}(j\omega)$ is shown[18] in Fig. 3. This function, as can readily be ascertained, crosses the real axis at the point -0.036, corresponding to $\omega=11.6$ sec^{-1}. It is thus clear that for values of A —the magnitude of the loop gain at zero frequency—exceeding $1/0.036=27.9$ the Nyquist graph will encircle the point -1 and the system will be unstable. One may alternatively examine the stability of this mechanism by use of Eq. (5), which, for the *critical* value of A, assumes the form

$$p^3+24p^2+135p+3240=0.$$

The roots of this algebraic equation are found to be $+11.6j$, $-11.6j$, and -24, the first two of which imply the presence of undamped oscillatory terms in the complementary function for the governing differential equation. If A were *larger* than 27.9, the constant term of the equation would be increased, the complex roots would have positive real parts, and instability would be implied.

It is also of interest to examine separate graphs of the magnitude and phase of $(1/A)\mathfrak{y}(j\omega)\mathfrak{z}(j\omega)$. These are shown in Fig. 4, in which a logarithmic frequency scale is used and the magnitude of the function is plotted on a decibel scale. It will be observed that the phase shift around the loop becomes $-\pi$ at approximately the same frequency as that for which the slope of the companion curve becomes -12 db/octave. This occurs at $\omega=11.6$ sec^{-1}, for which $\mathfrak{y}\mathfrak{z}/A$ assumes the value 0.036 corresponding to -28.9 db. Thus, in agreement with what we saw previously, we have an upper limit of 28.9 db to the loop gain at zero frequency and, if the gain is raised to this critical value, oscillations of frequency $\omega=11.6$ sec^{-1} will be sustained.

It will be noted that the upper limit of 27.9 for the loop gain of the mechanism described satisfies in a somewhat weak manner the condition (8a) earlier specified and, even if it were practicable to approach this value closely in adjusting a servomechanism of this design, the fidelity of response might be regarded as inadequate. In addition, if one attempted to minimize such objections by operating at a zero-frequency gain level close to the critical value, such as $A=24$, the presence of slowly damped transient oscillations could be very troublesome. Figure 7 shows the effect of

[18] As an inspection of Fig. 3 will suggest, a Nyquist graph may frequently be made more readable if for the larger gain values a logarithmic scale is used. This has been done, for example, by Peterson, Kreer and Ware (reference 9). It might be even more suitable to employ a hyperbolic scale, such as has been discussed by R. Legros, *Ann. de Physique* (Sér. 12) 1, 335 (1946).

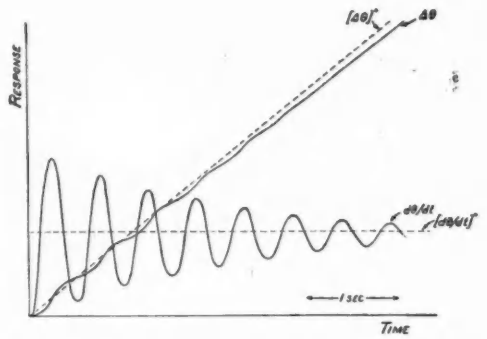

FIG. 7. Response of the servomechanism shown in Fig. 6(*a*) to a step-function input. The value of A, the magnitude of the loop gain at zero frequency, is taken as 24. Note the prominence of slowly damped transient oscillations.

such terms in the response[19] to an input that has the form of a step function.

With $A=24$, the *ultimate* response to a step function of height A_0 assumes the form

$$\Delta\theta \doteq (0.96t+0.0137)sA_0$$

and

$$d\theta/dt \doteq 0.96sA_0.$$

These particular solutions will result from direct consideration of the differential equations satisfied by the variables of the mechanism or may be obtained by reference to Eq. (11), which now assumes the form

$$(\mu/f)/(1+\mu\beta)=0.96+0.0137p-0.0089p^2-\cdots.$$

Similarly, the response to a ramp function of slope A_1 v/sec would be found to be

$$\Delta\theta \doteq (0.48t^2+0.0137t-0.0089)sA_1$$

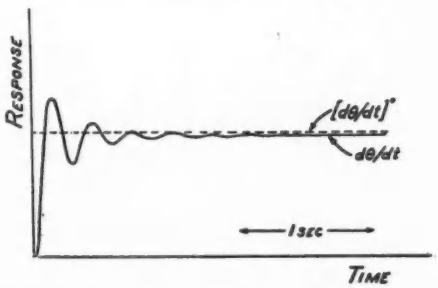

FIG. 8. Response of the servomechanism of Fig. 6(*a*) to a step-function input, using a feedback filter of the type shown in Fig. 6(*b*). The value of A', the loop gain at zero frequency, has been taken as 50.

[19] The methods of transient analysis can, of course, be very helpful in determining the solution, consistent with the appropriate initial conditions, of differential equations of the type with which we are concerned here.

and

$$d\theta/dt \doteq (0.96t+0.0137)sA_1,$$

whereas the *ideal* response would be

$$(d\theta/dt)^0 \doteq sA_1t \quad (\text{for } t\geqslant 0)$$

and

$$(\Delta\theta)^0 \doteq 0.5(sA_1)t^2.$$

In view of the unsatisfactory features of the servomechanism in the form considered, it is appropriate to attempt to modify the frequency-response curve in such a way that higher values of the zero-frequency loop gain are possible without instability. This can be done by the method suggested in SEC. 2 (*B*)(3), namely by introducing into the loop a filter that favors frequencies in the neighborhood of those for which the loop gain is dropping with undesirable rapidity. Such a filter, whose transfer function can be written $\eta(1+pT')/(1+pT)$ or $\eta(1+j\omega T')/(1+j\omega T)$, is readily constructed from resistances and capacitances in the manner indicated in Fig. 6(*b*). A filter of this type would appear to be useful if sufficient gain is available to overcome its attenuation at low and zero frequencies. If this filter is introduced into our circuit, with values 0.01 and 0.1 sec, respectively, for T and T', we have

$$\mu(j\omega)\beta(j\omega) = \frac{A'(1+j\omega/10)}{(1+j\omega)(1+j\omega/7)(1+j\omega/16)(1+j\omega/100)},$$

where $A'[\equiv\eta A]$ represents the magnitude of the loop gain at zero frequency. The coefficient η may be of the order of magnitude of 0.1.

Graphs of the magnitude and phase of $(1/A')\mu(j\omega)\beta(j\omega)$ for this new function are shown in Fig. 5 and are to be compared with those in Fig. 4, which applied in the absence of the filter. It will be observed that the phase angle now becomes $-\pi$ for a frequency in the neighborhood of $\omega=37$ sec^{-1} and that, corresponding to this frequency, we have a value of -43.0 db or an amplitude of 1/141 with respect to the value for $\omega=0$. It would appear, therefore, that by the addition of the filter described one could operate safely at a zero-frequency loop gain as high as 40 or 50, although possibly the transient effects which would be present with a gain margin this small would be considered unsatisfactory in certain applications.

The exact form of the response to various types

of input function will in general depend upon whether the filter is introduced into the amplifier circuit ($\mathfrak{y}$-circuit) or into the feedback path ($\mathfrak{\beta}$-circuit). Thus, for example, the velocity-error coefficient a_1 will assume a value that is different in these two cases. From a practical standpoint, however, it may be that the impedance and power level of the $\mathfrak{\beta}$-circuit is such as to make the insertion of the filter in this feedback portion of our mechanism by far the more convenient of the alternatives.

It is in any event clear that the incorporation of a filter of the type indicated would permit operation at a somewhat higher zero-frequency loop gain and therefore with an improved static error, while the transient oscillations resulting from various types of input function will be expected to be less troublesome. Figure 8 shows the result of the application of a step function to the servomechanism when the filter considered is inserted in the feedback circuit and the zero-frequency loop gain is given the value $A'=50$. Other types of filter could, of course, be devised for the further refinement of a servomechanism of the type considered here, but consideration of the various possibilities and their respective merits will not be undertaken in this paper.

It remains to be pointed out that if the device just considered is to be regarded as a *speed stabilizer*, we are then not concerned with changes in the input but in the effect of changes of $\mathfrak{y}$—specifically changes of $\mathfrak{y}(0)$. In accordance with the general result represented by Eq. (12b) we readily find that, with e_1 constant,

$$\delta(d\theta/dt)/(d\theta/dt)=(\delta\mu_0/\mu_0)/(1+\mu_0\beta_0),$$

where we have written μ_0 and β_0 for $\mathfrak{y}(0)$ and $\mathfrak{\beta}(0)$, respectively.

If changes in μ_0 due to possible changes in the flux density B of the motor field are being considered, as would be of interest with a field excited from an external source, μ_0 may vary inversely as B, provided $R\xi$ is small, and we can replace $\delta\mu_0/\mu_0$ by $-\delta B/B$. In such an application the effect of transient terms might not be of importance, since possibly a rapid variation of B would be precluded, and one could operate at a higher loop gain than would be permissible when the mechanism is subjected to rapid changes. In any case the presence of the factor $1+\mu_0\beta_0$ implies a stabilizing action which is due to the feedback and which becomes large for large values of the loop gain. It is necessary, however, that the $\mathfrak{\beta}$-circuit be designed to insure constancy of the transfer function for this latter circuit at the frequencies of interest, which requires in this illustration, constancy of β_0.

4. Summary and Discussion

In the present paper attention has been directed to the desirability of making some knowledge of servomechanism theory available to students receiving technical or scientific training. Such material would appear to be of value to numerous students who do not go on to advanced study in the fields of engineering or applied physics, and the author has attempted to outline a type of presentation that might be helpful to them. For the advanced student a more thorough study of control mechanisms, aided by the powerful methods of transient analysis, would, of course, be appropriate. A treatment of topics related to those discussed in the present paper, such as the study of multiloop and nonlinear systems, would certainly be of interest.

It is recognized that the discussion as given here is by no means complete and that an instructor teaching material of this nature will, of course, wish to amplify the discussion with examples of his own choosing, suitably timed to illustrate and clarify the theoretical treatment. It is acknowledged that, in some engineering curriculums, course work relating to the subject in question is available to advanced undergraduate students with specialized training. It is felt, however, that the introduction of this subject to a much larger body of students interested in a variety of technical fields merits the consideration of physics teachers.

The author would like to acknowledge his gratitude to several of his associates for helpful discussion and criticism of the manuscript for this paper.

If one advances confidently in the direction of his dreams, and endeavors to live the life which he has imagined, he will meet with a success unexpected in common hours.—THOREAU.

Caustics by Reflection in a Concave Spherical Reflecting Surface

G. F. HERRENDEN-HARKER
University College of South Wales and Monmouthshire, Cardiff, Wales

THE application of the laws of reflection to the individual rays emanating from a luminous point source and incident upon a reflecting surface determines a conjugate family of reflected rays which do not, in general, pass through a single point image. The only strictly stigmatic optical device is the plane mirror. Certain other optical instruments are stigmatic for selected positions of the point source; an example is the concave spherical mirror with the point source located at its center of curvature. Others again may be approximately stigmatic provided the incident pencil of rays is suitably restricted.

Despite the caution generally given in textbooks of optics that the diagrams are intentionally exaggerated in the interests of clarity, the beginning student is apt to acquire from them a false impression concerning the limitations of optical instruments in respect of stigmatism. The rational approach to the concept of stigmatism, basic in geometrical optics, implies a study of caustics, which assume their simplest forms when the optical system has an axis of symmetry and when the point source is located thereon.

Methods of tracing caustic curves with the aid of pins, often set as elementary laboratory exercises, afford only crude approximations to such curves since, though the consecutive rays defined by the pin punctures may be neighboring, they are by no means adjacent. Much more accurate representations of caustic curves can be constructed by plotting a series of points on the curves calculated from simply derived formulas. Appropriate formulas applicable to the reflection of rays from a point source on the axis of a concave spherical surface are deduced in this paper, and with the aid of them a number of typical caustic curves are drawn.

Let P (Fig. 1) be a point source of light and C the center of curvature of a concave spherical reflecting surface of radius r. The join of PC, defining the axis of the reflecting surface with respect to P, intersects this surface in M_0, the vertex of the reflecting surface corresponding to P. The circular arc M_0M represents a meridian section of the reflecting surface containing P.

Let PM be a ray in this plane incident at an angle i upon the mirror at M. Then, since MC is the normal to the reflecting surface at M, the ray reflected from M will lie in the plane PMC on the opposite side of the normal to the incident ray and inclined at an angle i to this normal.

Let the circular arc M_0M'' represent another meridian section of the spherical reflecting surface, M'' being the point that M would occupy if the figure were rotated through an angle about PCM_0 as axis. Then, the ray PM'' being likewise incident at an angle i upon the mirror at M'', and $M''C$ being the normal to the reflecting surface at M'', the ray reflected from M'' will lie in the plane $PM''C$ on the opposite side of the normal to the incident ray and inclined at an angle i to this normal. The pair of rays PM, PM'' reflected from M and M'', respectively, will accordingly intersect in the point Q, and this will be located upon the axis PCM_0.

The rotation envisaged need not be confined to a small angle since M'' will clearly be competent to act in a similar reflecting capacity in respect of rays originating from P, whatever position M'' occupies upon the circumference of center N and radius NM, where N is the foot of the perpendicular from M on the axis of rotation. If $PM \equiv u$, $CM \equiv r$, $QM \equiv v$ and angle $M_0CM \equiv \theta$, reference to Fig. 1 shows that from the triangle CPM,

$$u/\sin(\pi-\theta) = r/\sin(\theta-i),$$

and from the triangle CQM,

$$v/\sin\theta = r/\sin[\pi-(\theta+i)],$$

from which

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{r\sin\theta}[\sin(\theta-i)+\sin(\theta+i)],$$

that is,

$$(1/u)+(1/v)=2\cos i/r. \qquad (1)$$

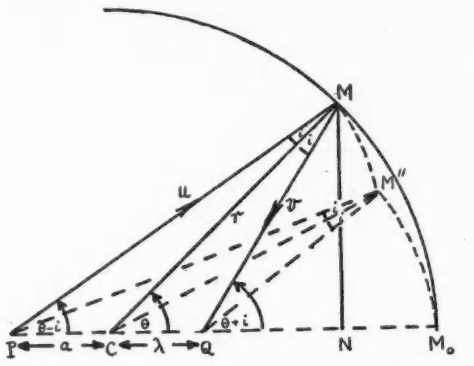

FIG. 1. Reflection of rays from a point source P on the axis of a concave spherical reflecting surface at points M and M'' on the circumference of a small circle of radius NM.

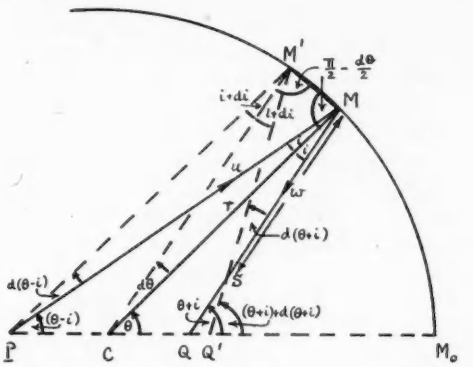

FIG. 2. Reflection of rays from a point source P on the axis of a concave spherical reflecting surface at points M and M' on a meridian through M_0.

In practice the reflected ray MQ corresponding to the incident ray PM is most conveniently fixed, when M is given, by locating Q. If $CQ \equiv \lambda$ and $PC \equiv a = r/\alpha$, say, so that P will be outside the spherical reflecting surface if $\alpha < 1$ and inside it if $\alpha > 1$, then from the triangle CPM,

$$a/\sin i = r/\sin(\theta - i),$$

that is,

$$\sin(\theta - i) = \alpha \sin i,$$

and from the triangle CQM,

$$\lambda/\sin i = r/\sin[\pi - (\theta + i)],$$
$$\sin(\theta + i) = (r/\lambda)\sin i;$$

hence, by subtraction,

$$2\cos\theta = (r/\lambda) - \alpha, \tag{2}$$

so that

$$\lambda = r/(\alpha + 2\cos\theta). \tag{3}$$

Similarly, by addition,

$$2\sin\theta = \left(\frac{r}{\lambda} + \alpha\right)\tan i. \tag{4}$$

Subtraction of Eq. (2) from Eq. (4) gives

$$(\sin\theta/\tan i) - \cos\theta = \alpha,$$

from which

$$\tan i = \sin\theta/(\alpha + \cos\theta).$$

Differentiation of this last expression gives

$$\sec^2 i \frac{di}{d\theta} = \frac{\cos\theta(\alpha + \cos\theta) + \sin^2\theta}{(\alpha + \cos\theta)^2},$$

from which

$$\frac{di}{d\theta} = \frac{\alpha\cos\theta + 1}{\alpha^2 + 2\alpha\cos\theta + 1}.$$

The maximum value of i thus corresponds to $\cos\theta = -1/\alpha$ and is given by

$$i_{\max} = \arctan(\alpha^2 - 1)^{-\frac{1}{2}} = \arccos\left(1 - \frac{1}{\alpha^2}\right)^{\frac{1}{2}}.$$

Moreover, as

$$\cos i = (1 + \tan^2 i)^{-\frac{1}{2}} = \frac{\alpha + \cos\theta}{(\alpha^2 + 2\alpha\cos\theta + 1)^{\frac{1}{2}}},$$

Eq. (1) can be written in terms of θ as

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}\frac{\alpha + \cos\theta}{(\alpha^2 + 2\alpha\cos\theta + 1)^{\frac{1}{2}}}. \tag{1'}$$

Further, since from the triangle CPM,

$$u/\sin(\pi - \theta) = a/\sin i,$$

that is,

$$\frac{1}{u} = \frac{\sin i}{a\sin\theta} = \frac{\alpha}{r(\alpha^2 + 2\alpha\cos\theta + 1)^{\frac{1}{2}}},$$

we have

$$\frac{1}{v} = \frac{\alpha + 2\cos\theta}{r(\alpha^2 + 2\alpha\cos\theta + 1)^{\frac{1}{2}}},$$

or

$$v = \frac{r(\alpha^2 + 2\alpha\cos\theta + 1)^{\frac{1}{2}}}{\alpha + 2\cos\theta}. \tag{5}$$

Let M' (Fig. 2) be a point on the meridian section M_0M of the spherical reflecting surface

in the vicinity of M such that the ray PM' is incident at an angle $i+\mathrm{d}i$ upon the mirror at M'. Then the ray reflected from M'—lying in the plane PMC and being inclined at an angle $i+\mathrm{d}i$ to $M'C$, the normal to the reflecting surface at M'—and the ray reflected from M will intersect in the point S.

Hence if $MS\equiv w$ and angle $MCM'=\mathrm{d}\theta$, reference to Fig. 2 shows that from the triangle CMM',

$$MM'/\sin\mathrm{d}\theta = r/\sin(\tfrac{1}{2}\pi-\tfrac{1}{2}\mathrm{d}\theta),$$

that is, $MM'=2r\sin\frac{1}{2}\mathrm{d}\theta$; from the triangle PMM',

$$\frac{MM'}{\sin\mathrm{d}(\theta-i)} = \frac{u}{\sin(\tfrac{1}{2}\pi-\tfrac{1}{2}\mathrm{d}\theta+i+\mathrm{d}i)},$$

that is,

$$\frac{1}{u} = \frac{1}{2r\sin\frac{1}{2}\mathrm{d}\theta}\cdot\frac{\sin\mathrm{d}(\theta-i)}{\cos(\frac{1}{2}\mathrm{d}\theta-\mathrm{d}i-i)},$$

and from the triangle SMM',

$$\frac{MM'}{\sin\mathrm{d}(\theta+i)} = \frac{w}{\sin(\tfrac{1}{2}\pi-\tfrac{1}{2}\mathrm{d}\theta-i-\mathrm{d}i)},$$

that is,

$$\frac{1}{w} = \frac{1}{2r\sin\frac{1}{2}\mathrm{d}\theta}\cdot\frac{\sin\mathrm{d}(\theta+i)}{\cos(\frac{1}{2}\mathrm{d}\theta+\mathrm{d}i+i)}.$$

Adding the last two equations and retaining only the first-order differential terms, one gets

$$\frac{1}{u}+\frac{1}{w} = \frac{1}{r\mathrm{d}\theta}\left[\frac{\mathrm{d}\theta-\mathrm{d}i}{\cos i+(\frac{1}{2}\mathrm{d}\theta-\mathrm{d}i)\sin i} + \frac{\mathrm{d}\theta+\mathrm{d}i}{\cos i-(\frac{1}{2}\mathrm{d}\theta+\mathrm{d}i)\sin i}\right]$$

$$= \frac{1}{r\mathrm{d}\theta}\,\frac{2\mathrm{d}\theta\cos i}{\cos^2 i-2\mathrm{d}i\cos i\sin i}$$

$$= \frac{2}{r\cos i(1-2\mathrm{d}i\tan i)},$$

so that in the limit when $\mathrm{d}i\to 0$,

$$(1/u)+(1/w) = 2/r\cos i, \tag{6}$$

or, in terms of θ,

$$\frac{1}{u}+\frac{1}{w} = \frac{2}{r}\,\frac{(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}{\alpha+\cos\theta}. \tag{6'}$$

Using the previously found expression for $1/u$ in terms of θ, we have

$$\frac{1}{w} = \frac{1}{r}\,\frac{\alpha^2+3\alpha\cos\theta+2}{(\alpha+\cos\theta)(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}},$$

that is,

$$w = \frac{r(\alpha+\cos\theta)(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}{\alpha^2+3\alpha\cos\theta+2}. \tag{7}$$

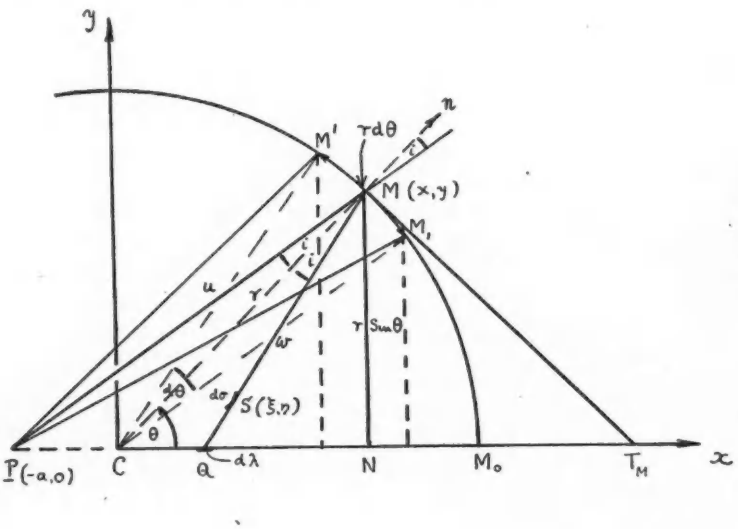

FIG. 3. Elements of the rectilinear and curvilinear arcs of a caustic curve due to reflection of rays from a point source P at points between M and M' on a meridian of a concave spherical reflecting surface through M_0.

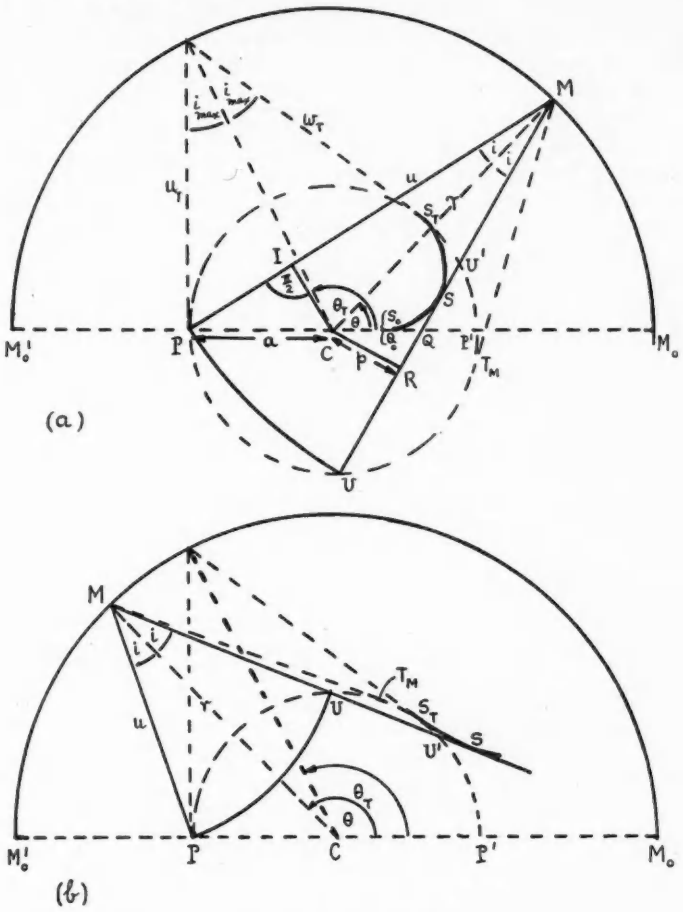

FIG. 4. Curvilinear arc of a caustic curve due to reflection of rays from a point source P at points on a meridian of a concave spherical reflecting surface through M_0: (*a*) $0<\theta<\frac{1}{2}\pi$, (*b*) $\frac{1}{2}\pi<\theta<\pi$.

The astigmatism of the pencil whose mean ray is specified by the angle θ and which is incident at an angle i on the spherical reflecting surface, being defined by $v-w$, will be given by

$$v-w=2r\frac{\sin^2\theta(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}{(\alpha+2\cos\theta)(\alpha^2+3\alpha\cos\theta+2)}. \quad (8)$$

The astigmatism vanishes for all values of α when $\sin^2\theta=0$, that is, when $\theta=0$ or π; in other words, the concave spherical mirror will be approximately stigmatic for all positions of the point source for sufficiently narrow pencils whose mean rays are directed along the axis and which are, therefore, incident quasinormally upon the reflecting surface. It also vanishes when $\alpha^2+2\alpha\cos\theta+1=0$; but, as this condition carries with it the implication that v and w are separately zero, it is irrelevant. When Eq. (8) is written in the form

$$v-w=2r\frac{\frac{\sin^2\theta}{\alpha^2}\left(1+\frac{2\cos\theta}{\alpha}+\frac{1}{\alpha^2}\right)^{\frac{1}{2}}}{\left(1+\frac{2\cos\theta}{\alpha}\right)\left(1+\frac{3\cos\theta}{\alpha}+\frac{2}{\alpha^2}\right)},$$

it is apparent that the astigmatism can vanish for all values of θ only if $\alpha=\infty$; in other words, the concave spherical mirror is strictly stigmatic only if the point source is located at its center.

Since θ enters into the expressions for v and w [Eqs. (5) and (7)] only in the form $\cos\theta$, their values will be unchanged when $-\theta$ is substituted

for $+\theta$. In other words, the section of the caustic surface by any meridian plane will consist of branches symmetrically disposed with respect to the axis. It is accordingly permissible to limit the discussion to values of θ included within the range 0 to π. As θ assumes in succession all values included within this range, the point Q will occupy successive points upon, and describe a rectilinear arc along, the line joining the point source P to the center C of the spherical mirror, while the point S will trace out a curvilinear arc in the meridian section defined by the plane containing P, C and M.

The complete caustic surface will be generated by rotating the plane loci traced out by Q and S, respectively, in any meridian plane through one whole revolution about PC as axis. It will consist of two sheets, one of which reduces to a linear segment (a cylinder of zero radius) extending over a portion of the axis, the other being the surface swept out by the curvilinear arc in the course of the rotation of the meridian plane containing it about the axis.

The position of the point source P will be specified in terms of its distance PC from the fixed point C (the center of the spherical reflecting surface of radius r) by $a=r/\alpha$.

The location of the point of incidence M on the spherical reflecting surface will be specified in terms of the angle θ between the line joining C to the vertex M_0 corresponding to P, and that joining C to M.

The locus of Q as θ varies will be given in terms of its distance $CQ[\equiv\lambda]$ from the fixed point C along the line joining C to M_0 by Eq. (3). Also, M being defined when θ is assigned, the locus of S as θ varies will be given in terms of its distance $MS[\equiv w]$ from M along the line joining M to Q by Eq. (7).

Such a procedure is better adapted to the plotting of points on the caustic curve than is one based on the evaluation of the Cartesian coordinates of the point S in terms of θ. Expressions for these coordinates are most readily derived by regarding the curvilinear branch of the caustic as the envelope of the family of reflected rays.

Referred to rectangular axes (Fig. 3) through C with the axis of abscissas along CM_0, the equation to the ray reflected from M—passing through the point $(r\cos\theta, r\sin\theta)$ and making an angle $\theta+i$ with the x axis—will be given by

$$y-r\sin\theta=\tan(\theta+i)(x-r\cos\theta),$$

that is,

$$y\cos(\theta+i)-x\sin(\theta+i)+r\sin i=0,$$

or, since

$$\tan i=\sin\theta/(\alpha+\cos\theta),$$

by

$$y(\alpha\cos\theta+\cos 2\theta)-x(\alpha\sin\theta+\sin 2\theta)+r\sin\theta=0.$$

Following the standard prescription we differentiate the last equation with respect to θ, obtaining

$$-y(\alpha\sin\theta+2\sin 2\theta)-x(\alpha\cos\theta+2\cos 2\theta)+r\cos\theta=0.$$

The solution of this pair of linear equations in x and y gives the Cartesian coordinates of the points S on the envelope in terms of the parameter θ. Denoting these coordinates by ξ and η, respectively, we have

$$\xi=r\frac{\alpha+3\cos\theta-2\cos^3\theta}{\alpha^2+3\alpha\cos\theta+2},\quad \eta=r\frac{2\sin^3\theta}{\alpha^2+3\alpha\cos\theta+2}. \tag{9}$$

An expression for the length of the curvilinear arc of the caustic is readily found. If, referred to the same rectangular axes as shown in Fig. 3, x, y are the coordinates of M, the point on the reflecting surface at which reflection takes place, and ξ, η are the coordinates of S, the point on the curvilinear arc of the caustic corresponding to

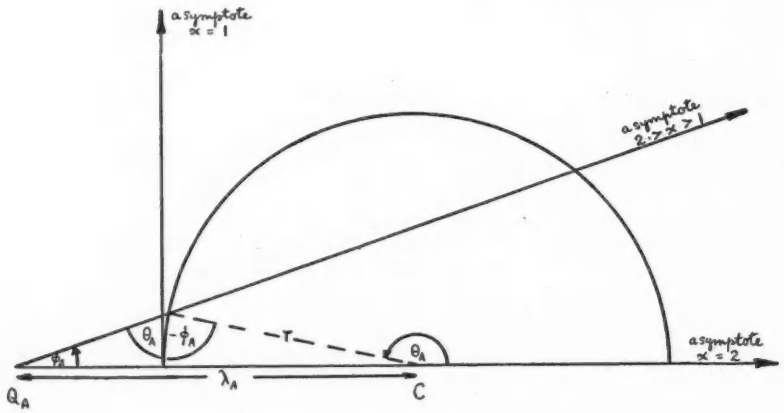

FIG. 5. Asymptotes to the curvilinear arc of a caustic curve due to reflection of rays from point sources on the axis at points on a meridian of a concave spherical reflecting surface, for $\alpha=2$, $2>\alpha>1$, and $\alpha=1$.

TABLE I. Values of $\cos\theta$, λ and w, for $r=10$ cm.

α	0	10/20	10/10	10/7	10/5	10/3	∞
$\cos\theta_0$	1	1	1	1	1	1	1
λ_0	5	4	3.33	2.92	2.5	1.87_6	0
w_0	5	6	6.67	7.08	7.5	8.12_5	10
$\cos\theta_V$	0.707	0.593	0.500	0.435	0.366	0.260	—
λ_V	7.07	5.93	5.00	4.35	3.66	2.60	—
w_V	3.54	4.73	5.77	6.53	7.34	8.51	—
$\cos\theta_{\pi/2}$	0	0	0	0	0	0	0
$\lambda_{\pi/2}$	∞	20	10	7	5	3	0
$w_{\pi/2}$	0	2.48	4.71	6.37	7.45	8.85	10
$\cos\theta_{M_0}$	0.5	0.25	0	-0.214_5	−0.5	—	—
λ_{M_0}	10	10	10	10	10	—	—
w_{M_0}	2.50	3.50	4.71	6.06	8.66	—	—
$\cos\theta_T$	—	—	−1	−0.7	−0.5	−0.3	0
λ_T	—	—	−10	348	10	3.66	0
w_T	—	—	0	7.14	8.66	9.54	10
$\cos\theta_H$	0	−0.25	−0.5	-0.714_5	−1	—	—
λ_H	∞	∞	∞	∞	∞	—	—
w_H	0	1.33	3.33	7.29	∞	—	—
$\cos\theta_A$	—	—	−1	−0.9429	−1	—	—
λ_A	—	—	−10	−21.9	∞	—	—
w_A	—	—	∞	∞	∞	—	—
$\cos\theta_\pi$	−1	−1	−1	−1	−1	−1	−1
λ_π	−5	−6.67	−10	−17.5	∞	7.5	0
w_π	5	3.33	0	−7.5	∞	17.5	10

reflection at M, then, since the coordinates of the point source P are $-a$, 0, we have

$$PM^2=(x+a)^2+y^2$$

and

$$SM^2=(x-\xi)^2+(y-\eta)^2;$$

or, remembering that $x^2+y^2=r^2$,

$$u^2=r^2+2xa+a^2$$

and

$$w^2=r^2-2(\xi x+\eta y)+(\xi^2+\eta^2).$$

Differentiation of these two relations gives $u\,du=a\,dx$ and $w\,dw=-(\xi\,dx+\eta\,dy)-(x-\xi)\,d\xi-(y-\eta)\,d\eta$.

The length $d\sigma$ of the element of the arc of the caustic in the vicinity of S will be given by $d\sigma^2=d\xi^2+d\eta^2$. Since SM is the tangent to this elementary arc at S, its direction cosines will be

$$\frac{d\xi}{d\sigma}\left(=\frac{x-\xi}{w}\right),\quad \frac{d\eta}{d\sigma}\left(=\frac{y-\eta}{w}\right),\quad 0.$$

Substitution in the second of the two relations obtained by differentiation gives

$$dw=-\frac{\xi\,dx+\eta\,dy}{w}-\left(\frac{d\xi}{d\sigma}d\xi+\frac{d\eta}{d\sigma}d\eta\right)$$
$$=-\frac{\xi\,dx+\eta\,dy}{w}-d\sigma.$$

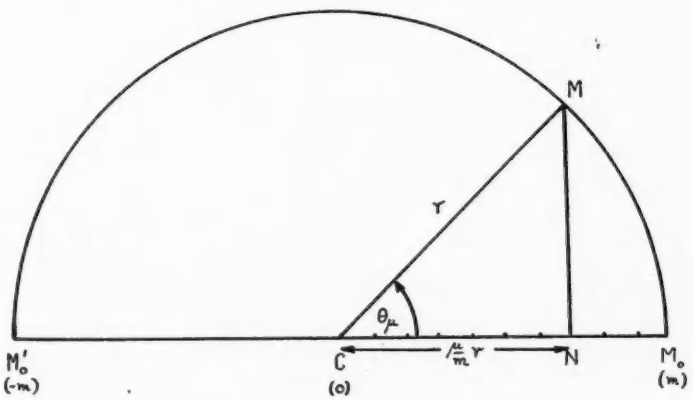

FIG. 6. Division of axis in preparation for plotting caustic curves.

TABLE II. Data for plotting the caustic curves.

μ \ κ	∞				20				10			
	λ		w		λ		w		λ		w	
10	(λ_0)	5.00	(w_0)	5.00	(λ_0)	4.00	(w_0)	6.00	(λ_0)	3.33	(w_0)	6.67
9		5.56		4.50		4.35		5.70		3.57		6.50
8		6.25		4.00		4.76		5.40		3.85		6.32
	(λ_V)	7.07	(w_V)	3.54								
7		7.14		3.50		5.26		5.08		4.17		6.15
6		8.33		3.00		5.88		4.75		4.54		5.96
					(λ_V)	5.93	(w_V)	4.73				
5	(λ_{M_0})	10.00	(w_{M_0})	2.50		6.67		4.41	(λ_V)	5.00	(w_V)	5.77
4		12.50		2.00		7.69		4.06		5.56		5.58
3		16.67		1.50		9.09		3.69		6.25		5.37
					(λ_{M_0})	10.00	(w_{M_0})	3.50				
2		25.00		1.00		11.11		3.31		7.14		5.16
1		50.00		0.50		14.29		2.90		8.33		4.94
0	$\binom{\lambda_{\pi/2}}{\lambda_H}$	∞	$\binom{w_{\pi/2}}{w_H}$	0	$(\lambda_{\pi/2})$	20.00	$(w_{\pi/2})$	2.48	$\binom{\lambda_{\pi/2}}{\lambda_{M_0}}$	10.00	$\binom{w_{\pi/2}}{w_{M_0}}$	4.71
−1						33.33		2.04		12.50		4.47
−2						100		1.58		16.67		4.22
					(λ_H)	∞	(w_H)	1.33				
−3						−100		1.08		25.00		3.94
−4						−33.33		0.56		50.00		3.65
−5						−20.00		0	(λ_H)	∞	(w_H)	3.33
−6										−50.00		2.98
−7										−25.00		2.58
−8										−16.67		2.11
−9										−12.50		1.49
−10									$\binom{\lambda_T}{\lambda_\pi}$	−10.00	$\binom{w_T}{w_\pi}$	0

μ \ κ	7				5				3			
	λ		w		λ		w		λ		w	
10	(λ_0)	2.92	(w_0)	7.08	(λ_0)	2.50	(w_0)	7.50	(λ_0)	1.87	(w_0)	8.13
9		3.10		6.98		2.63		7.46		1.95		8.15
8		3.30		6.89		2.78		7.43		2.03		8.18
7		3.53		6.79		2.94		7.39		2.11		8.22
6		3.80		6.69		3.12		7.37		2.21		8.26
5		4.12		6.59		3.33		7.35		2.31		8.32
	(λ_V)	4.35	(w_V)	6.53								
4		4.49		6.50		3.57		7.34		2.42		8.39
					(λ_V)	3.66	(w_V)	7.34				
3		4.93		6.41		3.85		7.34		2.54		8.47
									(λ_V)	2.60	(w_V)	8.51
2		5.47		6.32		4.17		7.36		2.68		8.57
1		6.14		6.24		4.54		7.39		2.83		8.69
0	$(\lambda_{\pi/2})$	7.00	$(w_{\pi/2})$	6.17	$(\lambda_{\pi/2})$	5.00	$(w_{\pi/2})$	7.45	$(\lambda_{\pi/2})$	3.00	$(w_{\pi/2})$	8.85
−1		8.14		6.10		5.56		7.55		3.19		9.03
−2		9.72		6.06		6.25		7.69		3.41		9.26
	(λ_{M_0})	10.00	(w_{M_0})	6.06								
−3		12.07		6.05		7.14		7.89	(λ_T)	3.66	(w_T)	9.54
−4		15.91		6.09		8.33		8.20		3.95		9.90
−5		23.33		6.21	$\binom{\lambda_{M_0}}{\lambda_T}$	10.00	$\binom{w_{M_0}}{w_T}$	8.66		4.29		10.35
−6		43.74		6.50		12.50		9.40		4.69		10.94
	(λ_T)	348	(w_T)	7.14								
−7		350		7.14		16.67		10.72		5.17		11.74
	(λ_H)	∞	(w_H)	7.29								
−8		−58.34		8.92		25.00		13.42		5.77		12.90
−9		−26.92		19.71		50.00		21.69		6.52		14.63
	(λ_A)	−21.88	(w_A)	∞	$\binom{\lambda_H}{\lambda_A}$	∞	$\binom{w_H}{w_A}$	∞				
−10	(λ_π)	−17.50	(w_π)	−7.50	λ_π		w_π		(λ_π)	7.50	(w_π)	17.50

The direction cosines of the incident ray PM being $(x+a)/u$, y/u, 0, those of the reflected ray MS being $(x-\xi)/w$, $(y-\eta)/w$, 0, and those of the tangent MT_M at M to the circular arc M_0M being dx/ds, dy/ds, 0, we have

$$\cos PMT_M = \cos\left(\frac{\pi}{2}+i\right) = -\sin i = \frac{(x+a)dx+ydy}{uds}$$

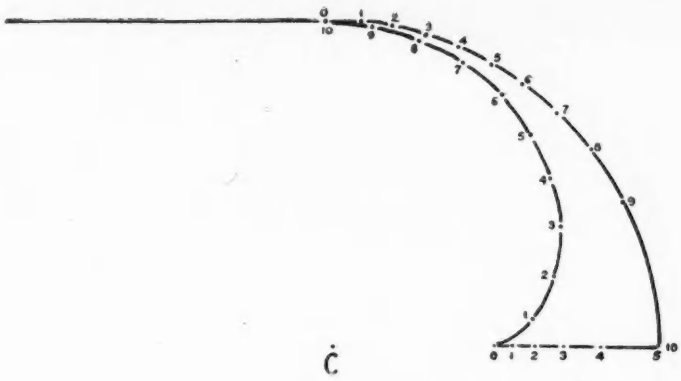

FIG. 7. Caustic curve in a meridian section of a concave spherical reflecting surface. Point source on the axis at infinity; $\kappa=\infty$, $\alpha=0$.

and

$$\cos SMT_M=\cos\left(\frac{\pi}{2}-i\right)=\sin i=\frac{(x-\xi)dx+(y-\eta)dy}{wds}.$$

Further, as the equation to the meridian section of the reflecting surface is $x^2+y^2=r^2$, we have $xdx+ydy=0$. Hence,

$$-\sin i=adx/uds=du/ds$$

and

$$\sin i=-\frac{\xi dx+\eta dy}{wds}=\frac{dw+d\sigma}{ds},$$

so that $d\sigma=-du-dw$, integration of which gives

$$\sigma=-(u+w)+\text{const.}$$

If the length of the curvilinear arc of the caustic is measured from the cusp on the axis corresponding to $\theta=0$ and for which $u=u_0$ and $w=w_0$, then

$$\sigma=(u_0+w_0)-(u+w)$$

But

$$u+w=r(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}\left(\frac{1}{\alpha}+\frac{\alpha+\cos\theta}{\alpha^2+3\alpha\cos\theta+2}\right)$$
$$=\frac{2r}{\alpha}\frac{(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}{\alpha^2+3\alpha\cos\theta+2},$$

so that

$$u_0+w_0=(u+w)_{\theta=0}=2r(\alpha+1)^2/\alpha(\alpha+2).$$

Hence

$$\sigma=\frac{2r}{\alpha}\left[\frac{(\alpha+1)^2}{\alpha+2}-\frac{(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}{\alpha^2+3\alpha\cos\theta+2}\right].$$

Thus, for example, as will appear later, when $\alpha=1$ the curvilinear arc of the caustic is a cardioid the half length of which will be given by $\sigma_{\substack{\alpha=1\\ \theta=\pi}}=(8/3)r$.

The light reflected from an arc M_1M' of length $rd\theta$ in the vicinity of M will pass through an arc of length

$$d\sigma=\frac{6r\sin\theta(\alpha\cos\theta+1)(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}{(\alpha^2+3\alpha\cos\theta+2)^2}d\theta$$

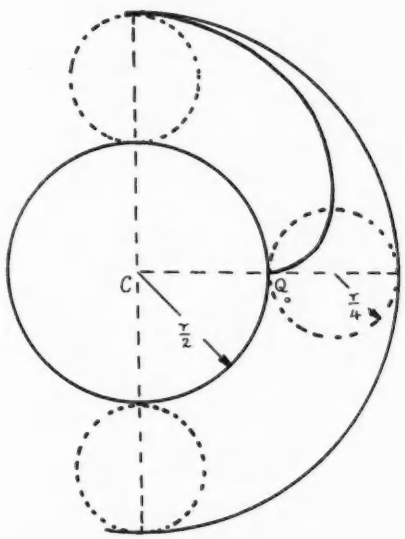

FIG. 7(a). The epicycloid generated by a point on the circumference of a circle of radius $r/4$ rolling externally on the circumference of a fixed circle of radius $r/2$ is the curvilinear arc of the caustic curve of Fig. 7.

in the vicinity of S and through a segment of length

$$d\lambda=\frac{2r\sin\theta}{(\alpha+2\cos\theta)^2}d\theta$$

in the vicinity of Q.

Assuming the point source at P to be of unit intensity and to radiate equally in all directions, the luminous energy impinging upon an elementary zone of the sphere of center C and radius r delimited by the pair of planes through M_1 and M' perpendicular to the axis PCM_0 and

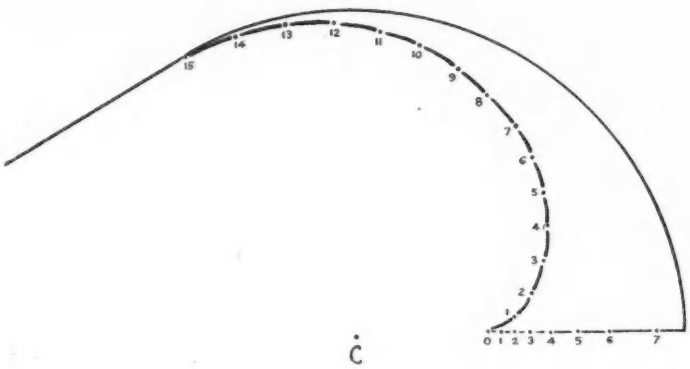

FIG. 8. Caustic curve in a meridian section of a concave spherical reflecting surface. Point source on the axis at a distance equal to the diameter from the center of the reflecting surface; $\kappa=20$, $\alpha=10/20$.

subtending a solid angle $d\Omega$ at P will be $dq=1\cdot d\Omega$. But

$$d\Omega=(rd\theta\cdot 2\pi r\sin\theta\cos i)/u^2,$$

or, since

$$\cos i=(\alpha+\cos\theta)/(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}$$

and

$$u=(r/\alpha)(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}},$$

$$dq=2\pi\alpha^2\frac{\sin\theta(\alpha+\cos\theta)}{(\alpha^2+2\alpha\cos\theta+1)^{\frac{3}{2}}}d\theta.$$

Hence the luminous energy per unit area of the elementary zone of the curvilinear branch of the caustic described by the rotation of the elementary curvilinear arc of length $d\sigma$ about the axis will be $dq/2\pi\eta d\sigma$ or, using the value for η from Eq. (9), will be given by

$$\frac{\alpha^2(\alpha+\cos\theta)(\alpha^2+3\alpha\cos\theta+2)^3}{12r^2\sin^3\theta(\alpha\cos\theta+1)(\alpha^2+2\alpha\cos\theta+1)^2}.$$

If θ is sufficiently small to permit us to write $\sin\theta=\theta$ and $\cos\theta=1$, this expression reduces to $\alpha^2(\alpha+2)^3/12r^2\theta^3(\alpha+1)$. It is apparent from the latter form of the expression that the concentration of luminous energy will be high in the vicinity of the cusp on the axis, approaching infinity as θ tends to zero, that is, at the cusp itself.

The luminous energy per unit area of the curvilinear branch of the caustic becomes infinite when $\alpha\cos\theta+1=0$, which corresponds, as will appear shortly, to the cusp located off the axis. Similarly, the luminous energy per unit length of the rectilinear branch of the caustic will be

$$\frac{dq}{d\lambda}=\frac{\pi\alpha^2(\alpha+\cos\theta)(\alpha+2\cos\theta)^2}{r(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}.$$

The complete discussion of the distribution of luminous intensity over the branches of the caustic is, however, beyond the province of geometrical optics.

Let $CR[\equiv p]$ be the length of the perpendicular from C on MQ, the ray reflected from M, as shown in Fig. 4(a). Then

$$p=r\sin i=r\sin\theta/(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}. \quad (10)$$

If CI is the perpendicular from C on PM, the ray incident at M, then $CI=CR=p$.

When $\theta=0$, $p=0$, in which case the corresponding values of λ, w and v become

$$\lambda_0=r/(\alpha+2), \quad w_0=v_0=r(\alpha+1)/(\alpha+2).$$

Since $w_0=v_0$, the two curvilinear branches of the caustic curve in a meridian plane, one above and the other below the axis, and the rectilinear branch of the caustic curve extending along the axis, all meet in a common point (S_0, Q_0) upon the axis. This point is a cusp and its distance from the center C is $r/(\alpha+2)$. Note also that $\lambda_0+w_0=\lambda_0+v_0=r$.

As θ increases, p likewise increases and assumes its maximum value a when the incident ray is normal to the axis, that is, for a value θ_T of θ given by $\cos\theta_T=-a/r$, or

$$\theta_T=\arccos(-1/\alpha),$$

which can be satisfied for values of $\alpha\geqslant 1$. This particular incident ray is the tangent at P to the circle of center C and radius $CP=a$. As the corresponding value of CR is then also equal to a, the reflected ray will likewise be tangential to this same circle at S_T.

When $\theta=\theta_T$ and the angle of incidence i assumes its maximum value, already shown to

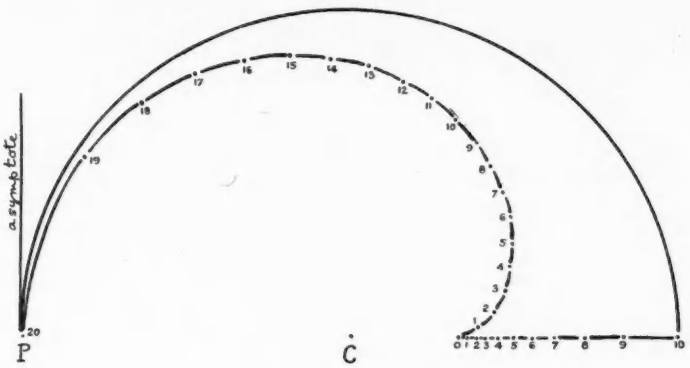

FIG. 9. Caustic curve in a meridian section of a concave spherical reflecting surface. Point source on the axis at a distance equal to the radius from the center of the reflecting surface; $\kappa=10$, $\alpha=10/10$.

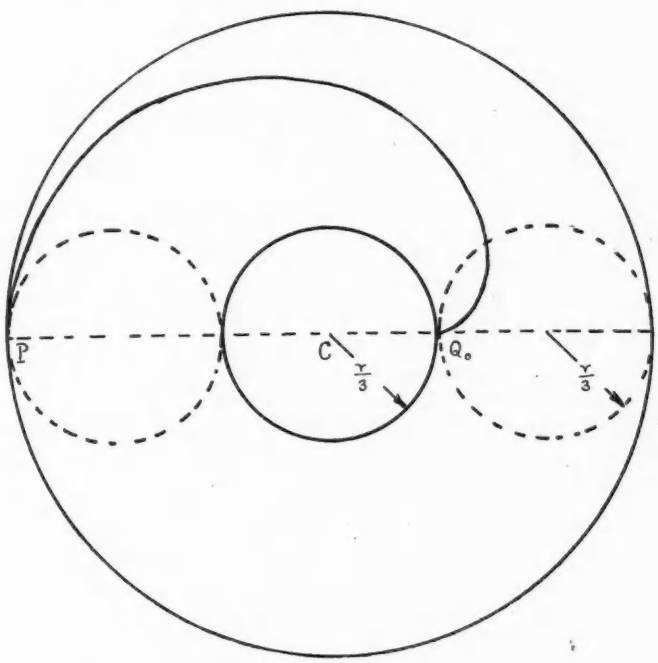

FIG. 9(*a*). The cardioid generated by a point on the circumference of a circle of radius $r/3$ rolling externally on the circumference of a fixed circle of radius $r/3$ is the curvilinear arc of the caustic curve of Fig. 9.

be arc cos $(1-1/\alpha^2)^{\frac{1}{2}}$, the corresponding values of λ and w are given by $\lambda_T=r\alpha/(\alpha^2-2)$,

$$w_T=(r/\alpha)(\alpha^2-1)^{\frac{1}{2}}=(r^2-a^2)^{\frac{1}{2}}=u_T.$$

If MT_M is the tangent from M to the circle of radius a and if the ray reflected from M cuts the circumference of this circle in U' and U, as is shown in Fig. 4, (*a*) and (*b*), then $MU'\cdot MU=MT_M^2$. Now

$$MT_M^2=r^2-a^2=(r^2/\alpha^2)(\alpha^2-1)$$

and

$$MU=MP=u=(r/\alpha)(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}},$$

so that

$$MU'=\frac{r}{\alpha}\frac{\alpha^2-1}{(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}.$$

The condition that $MS[=w]$ shall be greater than MU' is, therefore,

$$\frac{r(\alpha+\cos\theta)(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}}{\alpha^2+3\alpha\cos\theta+2} > -\frac{r}{\alpha}\frac{\alpha^2-1}{(\alpha^2+2\alpha\cos\theta+1)^{\frac{1}{2}}},$$

that is, $(\alpha\cos\theta+1)^2>0$, which will hold for all values of α and θ.

Over the range $\theta=0$ to $\theta=\arccos(-1/\alpha)$—$(1>\cos\theta>-1/\alpha)$—the curvilinear arc of the caustic situated above the axis will lie wholly within the semicircle PS_TP' of radius a. Over the range $\theta=\arccos(-1/\alpha)$ to $\theta=\pi-(-1/\alpha>\cos\theta>-1)$—the curvilinear arc of the caustic situated above the axis will lie wholly without the semicircle PS_TP' of radius a. These two portions of the curvilinear arc meet in a cusp S_T upon the circumference of this semicircle of radius a.

As θ continues to increase beyond the value $\arccos(-1/\alpha)$, p decreases from the value a and becomes zero again when $\theta=\pi$. The corresponding values of λ, w and v are given by $\lambda_\pi=r/(\alpha-2)$, $w_\pi=v_\pi=r(\alpha-1)/(\alpha-2)$. Since $w_\pi=v_\pi$, the two curvilinear branches of the caustic curve in a meridian plane, one above and the other below the axis, and the rectilinear branch of the caustic curve extending along the axis, all meet in a common point (S_π, Q_π) upon the axis. This point is a cusp, and its distance from the center C is $r/(\alpha-2)$. Note also that $w_\pi-\lambda_\pi=v_\pi-\lambda_\pi=r$. When $\theta=\pi/2$,

$$\lambda_{\pi/2}=a=r/\alpha, \quad w_{\pi/2}=r\alpha(\alpha^2+1)^{\frac{1}{2}}/\alpha^2+2,$$

and

$$v_{\pi/2}=(r/\alpha)(\alpha^2+1)^{\frac{1}{2}}.$$

The curvilinear branch of the caustic curve will have a vertical tangent if the ray reflected from M is perpendicular to the axis. This occurs when $\lambda=r\cos\theta$, that is, when

$$1/(\alpha+2\cos\theta)=\cos\theta,$$

or

$$\cos\theta=\frac{-\alpha\pm(\alpha^2+8)^{\frac{1}{2}}}{4}.$$

Since the relevant value of θ will necessarily lie between 0 and $\pi/2$, the appropriate value, θ_V, of θ will be given by $\theta_V=\arccos\frac{1}{4}[(\alpha^2+8)^{\frac{1}{2}}-\alpha]$. The curvilinear branch will always admit a vertical tangent.

The ray reflected from M will pass through the vertex M_0 if $\lambda=r$, that is, if $\alpha+2\cos\theta=1$, so that the corresponding value, θ_{M_0}, of θ will be given by $\theta_{M_0}=\arccos\frac{1}{2}(1-\alpha)$ and satisfied for values of $\alpha\leqslant 3$. The corresponding value of w will be given by $w_{M_0}=r(\alpha+1)^{\frac{1}{2}}/(4-\alpha)$.

The curvilinear branch of the caustic curve will have a horizontal tangent if the ray reflected from M is parallel to the axis. This occurs when $\lambda=\infty$, that is, for a value of θ given by

$$\theta_H=\arccos(-\tfrac{1}{2}\alpha)$$

and satisfied for values of $\alpha\leqslant 2$. The corresponding value of w will be given by $w_H=r\alpha/(4-\alpha^2)$. If $\alpha=2$, $\theta_H=\pi$ and $w_H=\infty$, so that the horizontal tangent coincides with the axis, which is an asymptote to the curvilinear branch.

In general, the asymptote to the curvilinear branch will be determined by the condition $w=\infty$, or $\alpha^2+3\alpha\cos\theta+2=0$, that is, will occur for a value of θ given by

$$\theta_A=\arccos\left(-\frac{\alpha^2+2}{3\alpha}\right)$$

and satisfied provided $2>\alpha>1$. The corresponding value of λ will be given by

$$\lambda_A=r\cdot 3\alpha/(\alpha^2-4)$$

and will be negative over the specified range of α. The asymptote passes through the point Q_A on the axis distant λ_A to the left of C. Assuming it

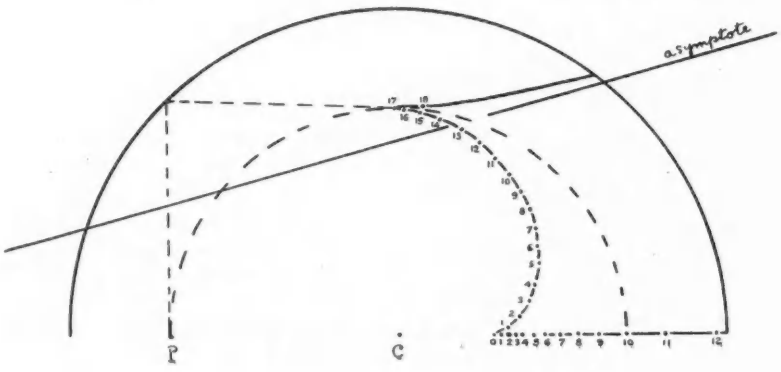

FIG. 10. Caustic curve in a meridian section of a concave spherical reflecting surface. Point source on the axis at a distance equal to 0.7 the radius from the center of the reflecting surface; $\kappa=7$, $\alpha=10/7$.

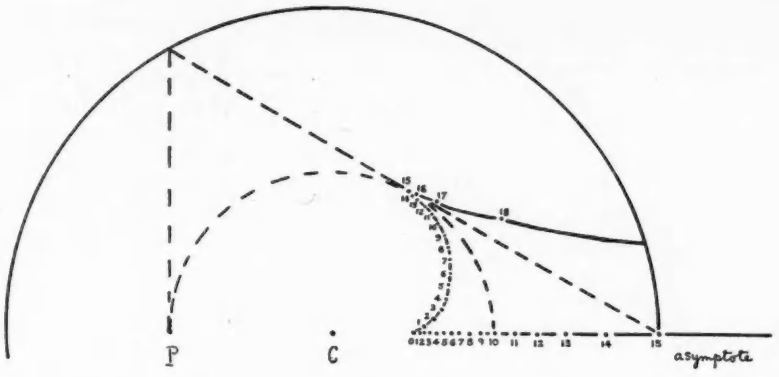

FIG. 11. Caustic curve in a meridian section of a concave spherical reflecting surface. Point source on the axis at a distance equal to 0.5 the radius from the center of the reflecting surface; $\kappa=5$, $\alpha=10/5$.

to make an angle ϕ_A with the axis, then, from Fig. 5, we have

$$r/\sin\phi_A=\lambda_A/\sin(\theta_A-\phi_A),$$

from which

$$\cot\phi_A=|3\alpha/(\alpha^2-4)|\ \mathrm{cosec}\theta_A+\cot\theta_A.$$

For example, if $\alpha=10/7$,

$$\cos\theta_A=-(\alpha^2+2)/3\alpha=-0.9429,$$

that is,

$$\mathrm{cosec}\theta_A=3.0031 \quad \text{and} \quad \cot\theta_A=-2.8318.$$

Hence

$$\cot\phi_A=2.188\times3.0031-2.8318=3.7390,$$

that is, $\phi_A=15°$.

Values of $\cos\theta$, λ and w corresponding to the various points of interest on the caustic as defined above are tabulated in Table I for the seven values of α listed and taking $r=10$ cm.

A convenient procedure to employ for the point-by-point plotting of the caustic curves is as follows. Having chosen a suitable value of the radius r of the spherical reflecting surface, we subdivide the diameter defining its axis into $2m$ equal parts, each of length r/m. Perpendiculars to the axis are drawn through the respective points of division to intersect the circumference of radius r. Then if N (Fig. 6) is the μth point of division to the right of C, so that $CN=\mu r/m$, and if the perpendicular to the axis through N intersects the circumference in M, the angle $MCM_0[=\theta_\mu]$ will be given by $\cos\theta_\mu=\mu/m$.

If these points of division are numbered from right to left as m(at M_0), $m-1$, $\cdots$, 0(C), -1, $\cdots$, $-m$(at M_0'), $\cos\theta$ will assume, in succession, the values 1, $(m-1)/m$, $\cdots$, $1/m$, 0, $-1/m$, $\cdots$, -1. Corresponding to reflection at M, the point Q on the rectilinear branch of the caustic will be fixed by its distance CQ along the axis from C, where

$$CQ=\lambda_\mu=\frac{r}{\alpha+2\cos\theta_\mu}=\frac{m}{m\alpha+2\mu}r.$$

The corresponding point S on the curvilinear branch of the caustic lies on the line joining M to Q at a distance from M given by

$$MS=w_\mu=\frac{r[\alpha+(\mu/m)][\alpha^2+2\alpha(\mu/m)+1]^{\frac{1}{2}}}{\alpha^2+3\alpha(\mu/m)+2}.$$

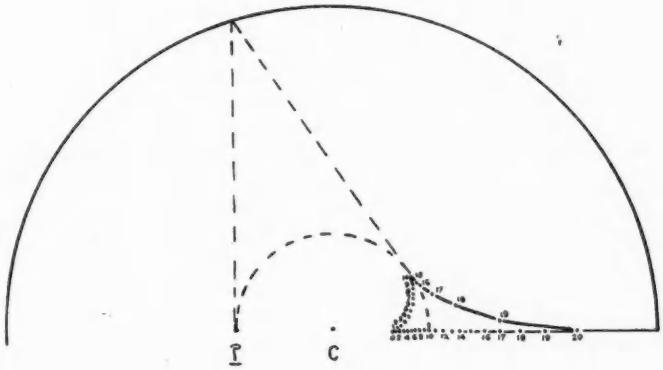

FIG. 12. Caustic curve in a meridian section of a concave spherical reflecting surface. Point source on the axis at a distance equal to 0.3 the radius from the center of the reflecting surface; $\kappa=3$, $\alpha=10/3$.

A slightly more compact form can be given to these expressions if the point source occupies one of the points of subdivision of the axis. If, for example, this is the κth to the left of C, so that $CP=\kappa r/m=a=r/\alpha$, then

$$\alpha=m/\kappa$$

and

$$\lambda_\mu=\frac{\kappa m}{m^2+2\kappa\mu}r,$$

while

$$w_\mu=\frac{(m^2+\kappa\mu)(m^2+2\kappa\mu+\kappa^2)^{\frac{1}{2}}}{m(m^2+3\kappa\mu+2\kappa^2)}r.$$

Taking $r=10$ cm and $m=10$, these reduce to

$$\lambda_\mu=50\kappa/(50+\kappa\mu) \text{ cm}$$

and

$$w_\mu=\frac{(100+\kappa\mu)(100+2\kappa\mu+\kappa^2)^{\frac{1}{2}}}{100+3\kappa\mu+2\kappa^2} \text{ cm}.$$

For the seven representative values of α listed in Table I the values of κ are, respectively, ∞, 20, 10, 7, 5, 3, 0.

The values of λ_μ and w_μ ($\mu=10, 9, \cdots, 0, \cdots, -10$) corresponding to the first six of these values of κ are contained in Table II.

The data listed in Table II have been plotted to give the caustic curves of Figs. 7 to 12. In Fig. 7 the point source is located on the axis at infinity. In this case the rays are incident parallel to the axis and the curvilinear arc of the caustic is the epicycloid generated, as is indicated in Fig. 7(a), by a point on the circumference of a circle of radius $r/4$ rolling externally on the circumference of a fixed circle of center C and radius $r/2$. The coordinates of the single cusp Q_0 on the axis are (5, 0 cm).

In Fig. 8 the point source is at a distance from the center of the spherical reflecting surface equal to its diameter. The coordinates of the single cusp on the axis are (4, 0 cm).

In Fig. 9 the point source is at one end of the diameter of the spherical reflecting surface. The curvilinear arc of the caustic is the cardioid generated, as indicated in Fig. 9(a), by a point on the circumference of a circle of radius $r/3$ rolling externally upon the circumference of a fixed circle of center C and radius $r/3$. The length of the arc Q_0P of the cardioid is four times the diameter of the fixed or rolling circle, in conformity with the result previously obtained. The coordinates of the single cusp Q_0 on the axis are (3.33, 0 cm).

In Fig. 10 the point source is 0.7 the radius, that is, 7 cm from the center of the spherical reflecting surface. The curvilinear branch exhibits three cusps. The cusp on the axis has coordinates (2.92, 0 cm). Those off the axis have coordinates (0.14, $\pm$7.00 cm). For, since $\cos\theta_T=-1/\alpha$, Eqs. (9) give

$$\xi_T=\frac{r}{\alpha^3}\frac{\alpha^4-3\alpha^2+2}{\alpha^2-1}, \quad \eta_T=\frac{2r}{\alpha^3}(\alpha^2-1)^{\frac{1}{2}},$$

and here $\alpha=10/7$ and $r=10$ cm. The asymptotes pass through the point ($-$21.88, 0 cm) and are inclined at an angle of 15° to the axis.

In Fig. 11 the point source is 5 cm from the center of the spherical reflecting surface. The coordinates of the cusp on the axis are (2.50, 0 cm). The coordinates of the cusps off the axis are (1.25, $\pm$4.33 cm). The asymptotes coincide with the axis.

In Fig. 12 the point source is 3 cm from the center of the spherical reflecting surface. The curvilinear arc exhibits four cusps. The coordinates of the pair on the axis are (1.87, 0 cm) and (7.50, 0 cm). The coordinates of the pair off the axis arc (2.46, $\pm$1.72 cm).

Note that, since

$$\lambda_{\substack{+\mu\\+\kappa}}=50\kappa/(50+\kappa\mu)=-\lambda_{\substack{-\mu\\-\kappa}}$$

and

$$\lambda_{\substack{-\mu\\+\kappa}}=50\kappa/(50-\kappa\mu)=-\lambda_{\substack{+\mu,\\-\kappa}}$$

while

$$w_{\substack{+\mu\\+\kappa}}=\frac{(100+\kappa\mu)(100+2\kappa\mu+\kappa^2)^{\frac{1}{2}}}{100+3\kappa\mu+2\kappa^2}=w_{\substack{-\mu\\-\kappa}}$$

and

$$w_{\substack{-\mu\\+\kappa}}=\frac{(100-\kappa\mu)(100-2\kappa\mu+\kappa^2)^{\frac{1}{2}}}{100-3\kappa\mu+2\kappa^2}=w_{\substack{+\mu,\\-\kappa}}$$

the caustic curves corresponding to negative values of κ (positions of the point source between C and M_0) will be the mirror images of the caustic curves corresponding to positive values of κ in the plane through C perpendicular to the axis.

Obvious is the most dangerous word in mathematics.—E. T. Bell.

On Probability Generating Functions

THEODORE JORGENSEN, JR.
University of Nebraska, Lincoln 8, Nebraska

THE method of calculating probabilities by the use of probability generating functions was first used by Laplace.[1] This method, sometimes mentioned by modern writers in the field of mathematical probability,[2–4] has recently come into use in the solution of problems arising in the study of nuclear chain reactions.[5] The purpose of this paper is to present the general aspects of probability generating functions through the discussion of a few interesting but simple examples.

Consider a trial, such as the throwing of a die, which results in an event with which is associated a certain score. Let p_j be the probability of an event happening with a score j. Let x be a variable, called a *hilfsbuchstabe*[2] (a help letter). The generating function for such a trial is then defined as the following sum:

$$P(x)=\sum p_j x^j,$$

where the summation extends over all the possible scores.

The simplest generating function is that for a trial (such as throwing a coin) with a probability p for a score of "unity" and a probability $q=1-p$ for a score "zero." It is $P(x)=qx^0+px^1=1+p(x-1)$. The generating function for the throwing of a single die is

$$P(x)=\tfrac{1}{6}[x^1+x^2+x^3+x^4+x^5+x^6].$$

Since the usual convention in probability theory makes the sum of the probabilities of an event unity, we have always $P(1)=1$. The expressions

$$dP(x)/dx=\sum jp_j x^{j-1}$$

and

$$d^2P(x)/dx^2=\sum j(j-1)p_j x^{j-2}$$

are useful since, when $x=1$, they represent $\bar{j}$ and $\bar{j^2}-\bar{j}$, respectively, where the bar above the letter indicates the mean value. The mean square deviation from the mean value $\langle j-\bar{j}\rangle_{Av}^2$, sometimes called the "dispersion,"[3] can be put in the form $\bar{j^2}-\bar{j}+\bar{j}-\bar{j}^2$. The dispersion is thus

$$\left.\frac{d^2P}{dx^2}\right|_{x=1}+\left.\frac{dP}{dx}\right|_{x=1}-\left(\left.\frac{dP}{dx}\right|_{x=1}\right)^2.$$

According to the theorem of compound probabilities,[3] the probability of a definite sequence of scores $s_1, s_2, s_3\cdots, s_m$ in m successive independent trials is the product of the probabilities of the various scores, $p_{s_1}p_{s_2}\cdots p_{s_m}$. If this product of probabilities is multiplied by the factor $x^{(s_1+s_2+\cdots+s_m)}$, the result can be arranged in this way:

$$p_{s_1}x^{s_1}p_{s_2}x^{s_2}p_{s_3}x^{s_3}\cdots p_{s_m}x^{s_m},$$

and this is the product of various terms in the generating functions of the single trials. The probability of obtaining a score $s_1+s_2+s_3+\cdots+s_m$ in this sequence is then one of the products $p_{s_1}p_{s_2}\cdots p_{s_m}x^{(s_1+s_2+\cdots+s_m)}$ in the expansion $Q(x)=\prod P_j(x)$, where $P_1(x)$ is the generating function of the first trial. However, the same total score,

$$V=\sum_1^m s_k,$$

may be obtained in other sequences and in other ways. According to the theorem of total probability,[3] the probabilities of obtaining a certain score V is the sum of the probabilities of the various exclusive ways of obtaining this score. This probability is the coefficient of x^V in the product $Q(x)=\prod P_j(x)$. This last expression is thus the generating function for m independent trials and is the product of the generating functions of the individual trials.

For the case in which the m events all have the same generating function, the generating function $Q(x)$ becomes $[P(x)]^m$. If $P(x)=q+px$, then

$$Q(x)=[q+px]^m=\sum_{j=0}^{j=m}\frac{m!q^{(m-j)}p^jx^j}{(m-j)!j!}$$

and the probability of having a score of N is the

[1] *Oeuvers de Laplace*, vol. 7, p. 1.
[2] E. Czuber, *Wahrscheinlickkeitsrechnung* (Leipzig, 1908).
[3] J. V. Uspensky, *Introduction to mathematical probability* (McGraw-Hill, 1937), p. 47.
[4] For an interesting application of generating functions, see R. H. Fowler, *Statistical mechanics* (Cambridge Univ. Press, 1929), p. 24.
[5] E. D. Courant and P. R. Wallace, *Phys. Rev.* **72**, 1038, 1947.

coefficient $m!q^{(m-N)}p^N/N!(m-N)!$ of x^N in this series. The mean value of the score is found by $dQ/dx|_{x=1}$ to be mp, and the dispersion is found by

$$\left.\frac{d^2Q}{dx^2}\right|_{x=1}+\left.\frac{dQ}{dx}\right|_{x=1}-\left(\left.\frac{dQ}{dx}\right|_{x=1}\right)^2$$

to be mqp. The standard deviation[3] is then $(mqp)^{\frac{1}{2}}$. Since the dispersion for the generating function $P(x)=q+px$ is qp, it is found that for a product of m such generating functions the dispersion is m times the dispersion qp. It is easily shown that the dispersion for the generating function $Q(x)=\prod P_j(x)$ is the sum of the deviations of the individual generating functions $P_j(x)$.

The generating function for the throwing of two dice is $[\frac{1}{6}(x^1+x^2+x^3+x^4+x^5+x^6)]^2$. This expands into the form $1/36[x^2+2x^3+3x^4+4x^5+5x^6+6x^7+5x^8+4x^9+3x^{10}+2x^{11}+x^{12}]$. The generating function for throwing a die and a coin, counting heads 1 and tails zero, is $[\frac{1}{6}(x^1+x^2+x^3+x^4+x^5+x^6)]\cdot[\frac{1}{2}x^0+\frac{1}{2}x^1]=\frac{1}{12}[x^1+2(x^2+x^3+x^4+x^5+x^6)+x^7]$.

The expression for the Poisson distribution may be obtained by the use of generating functions. Let random events happen at the average rate of M events in time interval T. Let $P(N,t)$ be the probability that exactly N events have happened since $t=0$. There can be N events at time $t+\Delta t$ if there were N events at time t and no event happened in the interval Δt, or if there were $N-1$ events at time t and one event happened in Δt, or if there were $N-2$ events at time t and two events happened in Δt, and so forth. The probability of an event happening in the interval Δt is $M\Delta t/T$, the probability of two events happening in Δt is $(M\Delta t/T)^2$. The probability of no event happening in Δt is

$$1-\frac{M\Delta t}{T}-\left[\frac{M\Delta t}{T}\right]^2-\cdots.$$

Since the various ways of having N events by time $t+\Delta t$ are mutually exclusive, the probability of having N events by $t+\Delta t$ is

$$P(N,t+\Delta t)=P(N,t)\left[1-\frac{M\Delta t}{T}-\left(\frac{M\Delta t}{T}\right)^2-\cdots\right]$$

$$+P(N-1,t)\left[\frac{M\Delta t}{T}\right]+P(N-2,t)\left[\frac{M\Delta t}{T}\right]^2+\cdots.$$

In the limit as Δt approaches zero this becomes the differential equation:

$$\frac{dP}{dt}=\frac{M}{T}[P(N-1,t)-P(N,t)].$$

To obtain an equation involving a generating function, multiply both sides of this equation by x^N and sum over all values of N. If $G(x,t)=\sum P(N,t)x^N$, then $xG(x,t)=\sum P(N-1,t)x^N$, and the differential equation becomes

$$\frac{\partial G}{\partial t}=\frac{M}{T}(x-1)G(x,t).$$

The solution of this differential equation which fits the boundary conditions is $G(x,t)=e^{(M/T)(x-1)t}$. The boundary conditions are $G(1,t)=1$ and $G(x,0)=1$. The first is the usual statement that the sum of all the mutually exclusive probabilities must be unity; the second is the statement that the probability of having a number N at $t=0$ different than zero is zero.

On putting $t=T$ the generating function becomes

$$G(x)=e^{M(x-1)}=\sum_0^\infty\frac{e^{-M}M^jx^j}{j!}.$$

The coefficient of x^N here is $e^{-M}M^N/N!$. This is the probability of N events happening in time T when the average is M. This is known as Poisson's distribution. By the method just described the standard deviation is found to be $M^{\frac{1}{2}}$.

The method of generating functions will now be applied to the problem of the disintegration of a sample of radioactive material. From experiment it is known that the fractional rate of disintegration for a large number of atoms is a constant. This is expressed by the equations $dN/dt=-\lambda N$ and $N=N_0e^{-\lambda t}$, where N_0 is the number of atoms at time $t=0$. Actually the disintegrations take place as discrete events, and the equation cannot describe accurately the number of atoms at any time in a sample. The experiment may be interpreted as determining the probability λ that a given atom will disintegrate in any particular unit of time, and $N(t)$ may then be considered to be the probable number of atoms at time t or the average number of atoms at time t in a very large number of similar samples. The

probability of an atom disintegrating in an interval Δt is then $N\lambda\Delta t$.

The problem at hand is that of determining the probability $P(N,t)$ of having any number of atoms N remaining after a time t. This can be done in the following way. Consider the probability $P(N,t+\Delta t)$. There can be N atoms at time $t+\Delta t$ if there were $N+1$ atoms at time t and only one disintegrated in the interval Δt or if there were N atoms at time t and no atom disintegrated in Δt. The probability of arriving at N atoms at time $t+\Delta t$ by the first process is the probability, $P(N+1,t)$, of having $N+1$ atoms at time t multiplied by the probability, $(N+1)\lambda\Delta t$, that one of these atoms disintegrates in Δt. This is $(N+1)\lambda\Delta t P(N+1,t)$. The probability of arriving at N atoms at time $t+\Delta t$ by the second process is the probability, $P(N,t)$, of having N atoms at time t multiplied by the probability, $(1-\lambda N\Delta t)$, that not one of these atoms does disintegrate in Δt. This is $(1-N\lambda\Delta t)P(N,t)$. The probability of having $N+2$ atoms at time t and two atoms disintegrating in time Δt will depend on $(\Delta t)^2$. Probabilities involving higher powers of Δt have been neglected.

By the theorem of total probability, $P(N,t+\Delta t)=(N+1)\lambda\Delta t P(N+1,t)+(1-N\lambda\Delta t)P(N,t)$, and $[P(N,t+\Delta t)-P(N,t)]/\Delta t=\lambda[(N+1)P(N+1,t)-NP(N,t)]$. In the limit as Δt approaches zero this becomes $(1/\lambda)(dP/dt)=[(N+1)P(N+1,t)-NP(N,t)]$. To obtain an equation involving generating functions, multiply both sides of this equation by x^N and sum over all possible values of N. Then $(1/\lambda)(\partial/\partial t)\sum P(N,t)x^N=\sum x^N[(N+1)P(N+1,t)-NP(N,t)]$. If $F(x,t)=\sum P(N,t)x^N$, then

$$\partial F/\partial x=\sum NP(N,t)x^{N-1}=\sum (N+1)P(N+1,t)x^N$$

and

$$x\partial F/\partial x=\sum NP(N,t)x^N.$$

The equation then becomes $(1/\lambda)(\partial F/\partial t)=(1-x)(\partial F/\partial x)$. The required generating function is the solution of this partial differential equation satisfying certain boundary conditions. The boundary conditions are determined when the number of atoms at $t=0$ is specified to be N_0 and the number at $t=\infty$ is specified to be zero; $F(x,0)=x^{N_0}$ by the first condition, and $F(x,\infty)=x^0=1$ by the second; $F(1,t)=1$ since all generating functions must equal one when $x=1$.

The required solution to the differential equation is found by the usual methods[6] to be

$$F(x,t)=[1-(1-x)e^{-\lambda t}]^{N_0}=\sum_{j=0}^{j=N_0}\frac{N_0!(1-e^{-\lambda t})^{(N_0-j)}e^{-j\lambda t}x^j}{(N_0-j)!\,j!}.$$

The mean value is $\bar{N}=N_0e^{-\lambda t}$, as should be expected, and the dispersion is $N_0e^{-\lambda t}[1-e^{-\lambda t}]=\bar{N}(N_0-\bar{N})/N_0$. This last quantity is zero at $t=0$ and zero at $t=\infty$, as it should be, since at these times the numbers of atoms are definitely N_0 and zero, respectively. The maximum dispersion occurs when the material is half gone. The standard deviation is then $0.5N_0^{\frac{1}{2}}$.

Another interesting case is that of a radioactive chain consisting, for simplicity, of a parent substance with a single daughter substance. Let A and B represent the numbers of atoms of each kind at any time t, and A_0 the number of parent atoms at time $t=0$. Let λ_1 and λ_2 be the disintegration constants of the parent atom and daughter atom, respectively. Let $P(A,B,t)$ be the probability that A and B atoms of the two kinds exist at time t. The parent-daughter relationship requires that a decrease in A by one increases B by one. Thus the probability, $P(A,B,t+\Delta t)$, of having exactly A and B atoms at time $t+\Delta t$ is the sum of three terms. First, there can be A and B atoms at time t and nothing happen in Δt; this probability is $[1-A\lambda_1\Delta t-B\lambda_2\Delta t]\cdot P(A,B,t)$. Second, there can be $A+1$ and $B-1$ atoms at time t and one parent atom disintegrate in Δt. This probability is $(A+1)\lambda_1\Delta t P(A+1,B-1,t)$. Finally, there can be A and $(B+1)$ atoms at time t and one daughter atom disintegrate in Δt. This probability is $(B+1)\lambda_2\Delta t P(A,B+1,t)$. In the limit as $\Delta t\to 0$ the equation becomes

$$dP/dt=-(\lambda_1A+\lambda_2B)P(A,B,t)+(A+1)\lambda_1P(A+1,B-1,t)+(B+1)\lambda_2P(A,B+1,t).$$

This equation is similar to that obtained before except that now the probabilities are functions of two numbers A and B. The generating function must then contain two variables; the foregoing

[6] A. R. Forsyth, *Differential equations* (Macmillan, 1921), pp. 274ff.

equation is therefore multiplied by $x^A y^B$ and summed. Let $F=\sum P(A,B,t)x^A y^B$. Then

$$\sum AP(A,B,t)x^A y^B = x\partial F/\partial x,$$
$$\sum BP(A,B,t)x^A y^B = y\partial F/\partial y,$$
$$\sum (A+1)P(A+1,B-1,t)x^A y^B = y\partial F/\partial x$$

and

$$\sum (B+1)P(A,B+1,t)x^A y^B = \partial F/\partial y.$$

Therefore the differential equation is

$$\partial F/\partial t = \lambda_1(y-x)\partial F/\partial x + \lambda_2(1-y)\partial F/\partial y.$$

The boundary conditions are $F(x,y,0)=x^{A_0}$, $F(x,y,\infty)=x^0=1$ and $F(1,1,t)=1$. The first equation states that the probability of having A_0 atoms at $t=0$ is certainty, the second states that the probability of having zero atoms at $t=\infty$ is certainty, and the third states that the sum of the probabilities is unity at all times.

The solution that fits the boundary condition is

$$F(x,y,t)=\left[1+(x-1)e^{-\lambda_1 t} + \frac{(y-1)}{\lambda_2-\lambda_1}(e^{-\lambda_1 t}-e^{-\lambda_2 t})\right]^{A_0}.$$

Now

$$\partial F/\partial x|_{x=y=1} = A_0 e^{-\lambda_1 t} = \bar{A},$$

and

$$\partial F/\partial y|_{x=y=1} = \frac{A_0\lambda_1}{(\lambda_2-\lambda_1)}(e^{-\lambda_1 t}-e^{-\lambda_2 t}) = \bar{B}.$$

Also

$$\left.\frac{\partial^2 F}{\partial x^2}\right|_{x=y=1} = A_0(A_0-1)e^{-2\lambda_1 t},$$

and

$$\left.\frac{\partial^2 F}{\partial y^2}\right|_{x=y=1} = A_0(A_0-1)(e^{-\lambda_1 t}-e^{-\lambda_2 t})^2\left(\frac{\lambda_1}{\lambda_2-\lambda_1}\right)^2.$$

The dispersions for A and B can be put in the form $\bar{A}(A_0-\bar{A})/A_0$ and $\bar{B}(A_0-\bar{B})/A_0$, as before.

The examples discussed so far are all concerned with events which happen independently of one another so that the probabilities of compound events are products of individual probabilities of the single events. Certain problems do not fall in this category. An example is given by the following game. One die is thrown. A second die is thrown as many times as the score shown on the first die. The score of the game is the sum of all the points turned up by the second die. What are the probabilities of the various scores? Here the probability of the score produced by the second die is contingent on what score is obtained on the first die. Such problems can be treated by generating functions.

A trial D_2 resulting in events E_2 which can occur with various scores has a generating function $P(x)=p_0x^0+p_1x^1+p_2x^2+\cdots$. Let the number of times trial D_2 occurs depend on the score of some other trial D_1 resulting in events E_1, having a generating function $Q(y)=q_0y^0+q_1y^1+q_2y^2+q_3y^3+\cdots$. If D_1 occurs with a score 0 (probability q_0) then event D_2 cannot occur; if D_1 occurs with a score 1 (probability q_1) then D_2 can occur once with various scores; and so on for other scores of D_1.

The probability of a certain total score k, obtained when the first score is j, is the product of the probability q_j of obtaining a score j in trial D_1 and the probability p_k of obtaining a score k in the j successive trials D_2. This is given by the coefficient of x^k in the product $q_j[P(x)]^j$. The total probability of a total score k is then the coefficient of x^k in the generating function $\sum q_j[P(x)]^j$. This is seen to be $Q(P(x))$.

As an example of this use of generating functions, consider the problem of the loss of counts in a counter for which the dead time is α, and for which counts occuring in the dead time do not lengthen the time in which the counter remains dead. Let the average number counted in a given time interval T be M; the generating function is then $R(y)=e^{M(y-1)}$ if the events being counted are assumed to be random. This last assumption is not strictly true since the dead time would prevent the counter from counting a larger positive deviation from the average number M. Assume that the events being counted occur at random with an average N in the given time interval T. The generating function for events occuring in a time interval α has been shown to be $Q(x)=e^{N(x-1)\alpha/T}$.

The number of times the counter is dead in an interval T is determined by the number of counts in that interval. The loss of a count is thus contingent on the occurrence of counts. The generating function for total counting loss is then

$$R(x)=\exp[M[Q(x)-1]] = \exp[M[e^{(\alpha N/T)(x-1)}-1]].$$

The mean value of the counting loss is $M\alpha N/T$. The counting rate plus the counting loss, $N=M+M\alpha N/T$, is the correct rate; thus

$$N=\frac{M}{1-(\alpha M/T)}.$$

So far all the scores associated with the various events have been positive. The case of a counter with random background is one in which it is convenient to give the background counts negative scores. If H and J are, respectively, the background and total counts in a given time interval T, then the generating functions are, respectively, $P(x)=\exp[H(x^{-1}-1)]$ and $Q(x)=\exp[J(x-1)]$. The generating function for the net counts in interval T is then

$$R(x)=\exp[J(x-1)]\exp[H(x^{-1}-1)].$$

The average count determined from this generating function is $J-H$, as it should be, while the dispersion is $J+H$. Since H and J are the dispersions for the generating functions $P(x)$ and $Q(x)$, this is an example of the addition of dispersions of individual generating functions.

These illustrative examples are presented to show the usefulness of the method of probability generating functions in applications in nuclear physics, at a level where a limited knowledge of this field of physics will not preclude an appreciation of the method.

The Limits of the Scientist's Responsibility*

WALTER C. MICHELS
Bryn Mawr College, Bryn Mawr, Pennsylvania

IN the period that has elapsed since Hiroshima, no change has been more marked than that which has taken place in the relationship of the scientist himself to the political and economic world in which he lives. Aside from the fact that the scientist has been increasingly called upon for advice, the public has become aware of him in a new way. It is true that a few marked figures in the scientific field, such as Einstein, have always captured the public imagination, but I can think of no period in the past when so large a number of scientific names have been known to the man in the street. Urey, Szilard, Oppenheimer, Smythe, Meitner, Condon, Compton, Bush, Bacher, Bohr—these no longer need an introductory comment when they are mentioned in the press. Considering the justified news coverage that the atomic bomb has received, this is not remarkable. But it is a sign of the times that civic groups, clubs, church organizations and schools are clamoring for scientists as speakers, whether or not those scientists have made any personal contributions to the release of nuclear energy; that nationally circulated periodicals—*The Saturday Evening Post*, *The Atlantic Monthly*, *Harpers*, *Social Progress*, *The Friend*, *Liberty*, to name only a few—are devoting space to articles written by scientists; that the Emergency Committee of Atomic Scientists can successfully engage in a campaign to raise two million dollars to be spent by scientists in educating the public on the implications of atomic energy.

As a group we must be gratified at this growth of public interest. At the very time when the majority of us recognized that the possible consequences of our researches forced us to step out of our laboratories and into the public arena, we found our fellow citizens ready to accept our opinions at, or even above, the value which we put upon them. It is true that the concrete results of this acceptance are still meager, but no one who has experienced the disappointingly slow progress that takes place in research can be greatly cast down on discovering that he cannot remake the world over night. The mere fact that we have been called into the council chamber, that we have been given a voice out of proportion to our numbers in shaping national and international policy, means that we can

* Paper delivered at The Pennsylvania Conference of College Physics Teachers, October 25, 1947.

exert greater influence on the world than ever before.

Gratification, however, is an insufficient response to public recognition. With our increased influence we have acquired increased responsibility, and the manner in which we discharge that responsibility will determine whether or not we deserve it. I here want to invite attention not to the details of the way in which we should face our increased task, which are, as the textbooks say, "beyond the scope of this course," but rather to the factors that we must keep constantly in mind as we face it.

Since it is in the species of scientist that John Doe has reposed his confidence, rather than in individuals of that species, let us examine the common characteristics of those engaged in scientific pursuits. We may thus see in what ways they are qualified as leaders of opinion and, on the other hand, in what ways they must be cautious in exercising leadership.

By the term *scientist* I shall mean an individual who has been subject for a period of some years to the discipline inherent in the observational-theoretical-experimental method of Galileo, Newton and Darwin, and who has subsequently devoted a major portion of his time and interest either to the search for fundamental knowledge within the framework of that method or to the dissemination of such knowledge. I purposely exclude the engineer, whose interest in the applications of scientific knowledge to the practical ends of life give him a quite different point of view from that of the scientist. In what ways do the habits of thought of the scientist differ from those of his fellow citizens?

Probably the most pronounced of the scientist's characteristics, despite a common belief to the contrary, is his imagination. Since he lives, during a large share of his waking hours, in an invented world in which generalizations and concepts are at least as real as the direct evidence of his senses, he must have imagination or die. The daring imagination that allowed Newton to conceive of a gravitational field extending to the outermost reaches of space, or that caused Anderson to connect a pair of curved paths in his cloud chamber with the simultaneous creation of a positive and a negative electron, is in no way more limited than that of the poet who draws upon mental imagery to create a feeling of emotion. Because of this continual use of his imaginative faculties, the scientist is far more ready than most of his fellow men to visualize a political and economic world in which radical changes have taken place. He can temporarily transfer his life to that world just as he moves himself daily into a world in which neutrinos carry momentum and energy, or genes serve to transmit characteristics from father to son. We have all experienced, I am sure, the conflict that occurs when a scientist, discussing a problem with his non-scientific friends, insists on basing the discussion on a set of suppositions that are to him a possible basis for an objective world, but are intellectually or emotionally repugnant to the others of the company.

The history of science and the nature of the scientist's work have forced him to balance his imagination with an intellectual caution that appears to most outsiders to be extreme. We have all heard the story of the scientist and the artist who went on a stroll together. After walking for some time they sat on the edge of a field, enjoying the view about them. A sheep came through the field and the artist remarked, "Ah, there goes a shorn sheep." To which the scientist replied, "No, there goes a sheep shorn at least on this side." The development of science has exhibited so many errors of observation and reasoning that the scientist has learned to be critical of both his perceptions and his reasoning. In particular, he distrusts inductive processes and analogies until such time as the conclusions reached by these means have been carefully checked by deduction and by comparison with observation. Since so much of political argument is based on the broadest of intuitive generalizations, this attitude of caution often makes the scientist intolerant of everything outside his limited field. He sometimes fails to recognize that the pressure of events may force decisions for which no really sound basis can be found. The recent war, by bringing scientists and military men together to so great an extent, emphasized the conflict between the two traditional attitudes of "Let's find the facts before we go ahead" and "A bad decision is better than no decision."

The struggle between the antithetical qualities of imagination and caution might be expected to lead their possessor into schizophrenia, yet I believe that we show no more tendency in this direction than do others. The scientist avoids this danger by limiting his professional activities to relatively simple problems, problems which have been so delimited that they can be resolved in reasonable time despite the strong forces pulling in opposite directions. The essence of experimental technic lies in the control of all but one or two of the possible independent variables. This amounts to the replacement of the complicated problem found in nature by a similar but somewhat artificial problem, to which the methods that have been found profitable in the past can be applied. After a considerable number of these simple problems have been solved, one can sometimes synthesize the natural situation from them, but this is a secondary role of science. This process of simplification—or even over-simplification—is one without which the growth of science would have been impossible, but its limitations frequently become apparent. The development of gestalt psychology and the introduction of exchange forces in atomic physics are typical of the occasional dissatisfaction of the scientist with his previously satisfactory methods. He is sometimes forced to the unpleasant conclusion that the whole is greater than the sum of its parts.

It is also his habit of caution that gives the scientist his reputation for pedantry and causes many laymen to look upon science as something esoteric. Because intellectual caution demands, as a necessary corollary, an exactness of both statement and thought, the scientist is necessarily something of a semanticist, something of a logician, something of a mathematician. His insistence on exact definition, on precise categorization, on the unambiguous description inherent in the mathematical statement of a theory are at wide variance with the purposely vague and loosely woven arguments that characterize much of our political activity. This exactness is one of the scientist's most valuable possessions, but he pays a price for it, since it also forces him either to limit his considerations to relatively simple problems or to treat the more complicated ones in a way that is foreign to his professional activities. It is easy to write a differential equation connecting the force on a planet with its motion, if we can treat the planet as a point mass and neglect all of its other properties; it is impossible in our present state of knowledge to express in a similar way the complicated interactions which take place among the several billion human inhabitants of that planet.

Because his imagination is broad enough to allow the consideration of almost all possible ideas and because his caution will allow the acceptance of few, the scientist is necessarily a believer in free speech. He knows that, in scientific circles, the talking of absurdities can lead to no harm other than some waste of time, since each idea will be tested for soundness against the stones of rigor and experiment. On the other hand, he has seen apparent absurdities turned into accepted fact. Since he is not, within his own field, susceptible to spellbinding, he has no reason to fear free speech, and since he recognizes that his critical judgment must apply to himself as well as to others, he does not become hysterical when a colleague in a professional meeting dares to challenge his ideas. Scientific arguments are sometimes vitriolic, but when they become passionate they lose their scientific character and might better be transferred to a Congressional committee or to the United Nations rostrum.

The same factors that give the scientist a hearty respect for free speech make him a poor democrat. Scientific matters are never settled by majority rule. Unanimity, or at least essential unanimity, is required for the acceptance of an observation as a fact or of a theory as sound. Consequently, scientific controversies are not resolved by compromise. One must either convince his opponents completely or be convinced completely by them. If neither of these states can be reached, it is far better that the controversy continue. How different this is from the parliamentary procedure in which one is faced with either action or chaos and in which a compromise, often emotional rather than intellectual, is the only possible solution. If our training and experience have done anything for us, they have made us into open minded individualists—persons who are open to conviction,

but who will maintain the rights of themselves and others to present what they consider to be facts or sound logic even when these are in opposition to the most firmly planted of traditional and accepted ideas.

How do these characteristics, which are salient ones of scientists in general, bear upon the responsibility of the individual scientist toward society?

I suppose we can all agree that every citizen in a democracy has a double duty: first, to improve the material, intellectual, moral or other standards of the community by doing his own job to the best of his ability; and second, to participate in government to a degree that varies with the contribution he can make. At the least, he should keep himself sufficiently informed to cast his vote intelligently; at the most, he may be in honor bound to lay aside his business or professional activities in favor of a position in government. The latter course becomes desirable only if the special abilities of the individual are such that the contributions which he can make in this way are greater than those which he can make in his own job. The same criterion applies to the intermediate state of affairs in which a man finds himself when he possesses special knowledge that has a bearing on political issues, knowledge which he feels should be shared with other citizens. If that knowledge is of real importance, if its dissemination will lead to better political decisions, its possessor has a heavy responsibility. The cost, in time and energy, of educating a citizenry which frequently asks for nothing more than to be left alone is high. Many of those who try must, at times, wonder whether the game is worth the candle.

It is this possession of special knowledge that places the scientist in his present dilemma. He sees not only the terrific potentialities of atomic warfare, but the general danger which is being created by an accelerated technical advance geared to a slowly moving social system. He is in a position to place both upper and lower limits on the probable achievements in nuclear science in the near future, and therefore to form some estimates of its future effect both on the art of warfare and on the economics of peace. Similarly, he knows something of the role that bacteria and hormones play in the life process and so can anticipate the possible effects of existing biological knowledge on the related but antagonistic problems of biological warfare and public health control. The public has a right to these facts, and it is the clear duty of every scientist to take every practicable step to drive them home. We can take pride in the work that has already been done in this connection by individual scientists and by such groups as the Federation of American Scientists and the Emergency Committee of Atomic Scientists. The task is still, however, far from finished, and it is therefore easy to establish the minimum civic duty of the scientist. None of us has a right to lock himself behind his laboratory door until the possible consequences of a misuse of existing knowledge have been made clear to every mentally normal man and woman in the civilized world.

If the lower limit of the scientist's responsibility is easy to define, the same cannot be said of the upper limit. We all know that the establishment of basic facts is not enough either in science or in politics. In either case, the facts must be correlated and interpreted. Further, in the political sphere, one must arrive at a course of action which can be expected to lead to desirable objectives and which is consistent with the known facts. What can the scientist contribute to this phase of the political problem?

It is natural enough that each of us, when he faces a problem outside of his own field, should carry over to that problem the habits of thought which characterize his professional activities. So the scientist, forced to consider an economic or political situation, is likely to start by applying to it the same rigorous and exact reasoning that he uses in science. It does not take him long to discover that most of what he reads will not bear this type of criticism. The definitions are often vague, the reasoning is frequently loose. When he seeks the reason for these breaches of good logic, he recognizes that they are due to the complexity of the problem under consideration. At this point he probably does one of two things. He either decides that the job in hand is far too complicated to be amenable to scientific methods or he starts looking for the really relevant factors,

in order that he may reduce the task to one that can be reasonably handled.

If the complexity of the problem baffles him, his course of action usually depends on the strength of his social conscience. He may throw up his hands in despair, retire to his laboratory and let the rest of the world go by. If his emotions will not allow this, he is likely to join the *mea culpa* school, cry out that the results of his work are ruining civilization, cover himself with sackcloth and ashes, and set out as a politician to save the world by an emotional appeal. With either of these courses I have little sympathy, for the first merely refuses to recognize a responsibility, while the second ruins the scientist and puts him to work at a task that others can usually perform better than he.

If he adopts the more consistent policy of simplifying the problem, he is equally likely to be misled. The scientist has seldom had the time to acquaint himself with more than a smattering of economics, politics, and other social studies. His judgment as to which factors are important is therefore likely to be warped and, in his attempt to simplify the problem, he will probably oversimplify it in a totally illegitimate way. The resultant oversimplified problem is one that appeals to him, for he has reduced it to one which can be managed by his own technics. At the end of a week he begins to feel pity for the countless people who have failed to see how easily an application of the scientific method resolves what has appeared to be an impossible conflict. He then sets out to convince the world, by purely logical arguments, that his solution is the only possible one. He is so convinced of the soundness of his arguments that he has no patience with those who do not agree with him, and he certainly is not willing to accept compromise. He either acquires a superman attitude or makes himself absurd. In either case he is likely to produce the exactly opposite effect from what he intended.

At this point I may be accused of exaggeration, but an examination of some of our past efforts will show how our tendency to oversimplify can lead us astray. About two years ago, many of us became very worried about the dangers of atomic warfare. We recognized those dangers to be so great and the possibilities of military defense to be so small that another war seemed unthinkable. Since we had reached this conclusion almost unanimously, it was easy for us to decide that all that we needed to do was to make the facts clear to others and war would be abolished. What we ignored here was a political inertia that renders impossible sudden violent changes in our institutions. This inertia, of which those more experienced in molding opinion could have warned us, has allowed the world to become as afraid as we but has prevented it from reaching the remedy that seemed so clear to us. Hence one of our main accomplishments has been to encourage the school of thought, led by people such as George Earle, that calls for preventive atomic war. We, who recognize the folly of this school, have encouraged it by the fear psychology that we built up in our ignorance.

In other words, the scientist is likely to be led astray regardless of what course of action he chooses. Is he, then, to say, "Everything that I do may help lead the world into a worse mess than it is in already, so I shall return to my laboratory and take only the minimum responsibility of informing my fellow citizens of the technical facts involved, leaving the conclusions to them"? To that my answer is an unequivocal "No." The scientist, even though he may not be equipped to be the savior of mankind, must remember that he has definite qualifications in his special knowledge, in his practice in rigor, in his insistence on facts, and in his trained critical faculties. At the same time, he must concede that he has his limitations and that others also have contributions to make from the methods used in their fields of activity, however faulty those methods may be when applied to a purely scientific problem. The questions that plague us are too complex to be answered by anything short of the collaboration of the best available minds. Their answers will not be found by scientists alone, whether those scientists act individually or through their own organizations. The hope lies, rather, in groups which include a wide representation of varying experiences and talents. Probably the most important role of the scientific organizations is to act as condensation nuclei for these broader associations. This action, if it is to be successful, must be carried on in a spirit of combining forces to seek an answer, not of persuading the others that we already know the answer and merely want their help in the final engineering job of its application.

If the type of cooperation that I suggest is to be used, the scientist will find himself in a some-

what peculiar position. In the discussions within the group he can remain completely a scientist and can give free rein to his capacity for destructive criticism. When he presents the conclusions of the group to those outside of it, he will undoubtedly find himself departing from a purely scientific approach. That this should happen is no more or less than a consequence of the fact that he is both a scientist and a citizen. The danger here lies in the difficulty of making clear to an audience the capacity in which he is acting. He may, for example, point out on scientific grounds that there is no foreseeable effective defense against atomic and biological warfare and then go on to propose some form of world government as the best means to save civilization. Even if he is very careful to point out that the final proposal is based on less rigorous arguments than the premise on which it rests, he is likely to find himself quoted as saying that scientists have proved that a particular form of world government is the only way to survival. No immediate harm results, but the failure to achieve world government in the next year or so may bring about a state of frustration and despair in those whom he has unintentionally misled—a state in which there is no desire for further search for a remedy.

This danger is one that cannot be completely removed, but it can be minimized if we make every effort to avoid the use of scientific prestige to strengthen nonscientific arguments. As scientists, we can present only those conclusions to which the rule of essential unanimity applies, that is, only those with which an overwhelming majority of our colleagues agree in detail. It would be comforting if we could here take a lesson from the Roman Catholic Church, and could speak *ex cathedra* on such subjects. In the absence of such a label, we can only use every possible care to separate the scientist and the citizen at the presentation level while we thoroughly mix the two at the thought level. This is not easy, but an application of the same judgment that we use in our laboratories and classrooms should make it at least partially successful.

Science is rapidly destroying, or at least overriding and overruling, the old boundaries. We are fast becoming neighbors of one another with far-reaching implications for old concepts and practices like war, race, nationality and the like. Neighborliness, the "good neighbor," are not only ethical concepts but are rapidly becoming economic, political and international concepts and standards of human behavior which we violate at our peril. This may be a hard saying, but war is the hardest of all teachers and its nameless sufferings and sacrifices in our generation are at last driving home to us the inner meaning of what religion has always taught us and science is now rediscovering for us.—JAN CHRISTIAAN SMUTS.

The Scientist and the Common Good*

ERIC D. MCCORMACK, O.S.B.
St. Vincent College, Latrobe, Pennsylvania

WILLIAM JAMES' happy phrase that to an infant the world is a big, buzzing, blooming confusion describes the beginning of our knowledge sufficiently well. Progress from that stage consists in distinguishing one thing from another, and in gradually discovering larger and larger patterns in which things thus discovered belong. The aim of science or knowledge would accordingly seem to be knowledge of the entire universe, and of the order in which its contents are arranged. Thus, it seems of advantage to begin by consulting some men of science on the universe and its order.

> I believe in God, the God of Spinoza, who reveals himself in the orderly harmony of the Universe. . . . I believe that intelligence is manifested throughout all nature. . . . The basis of all scientific work is the conviction that the world is an ordered and comprehensible entity—not a thing of chance.[1]
>
> Certain it is that a conviction, akin to religious feeling, of the rationality or intelligibility of the world lies behind all scientific work of a higher order.[2]
>
> . . . There can be no living science unless there is a widespread instinctive conviction in the existence of an *Order of Things*, and in particular, of an Order of Nature.[3]
>
> That there is something of a nonphysical nature which controls the action of the atom is quite in the air today. . . . It is permissible to postulate a fundamental unity or order in which all living things, as well as all nonliving things, are controlled by something which approaches consciousness.[4]

Here we have an acknowledgment of a real order in the universe, an order existing prior to our knowledge. Let us pass over the metaphysical problem of its origin and ask these men how we become aware of this order. One of them finds it by a feeling, another by an instinct; the third more cautiously calls it a postulate. All appear to agree that the principle itself of order is not to be found by the methods of empirical science, and at least two seem to suggest that this principle is to be found by some other means than reason. Reason is here made coextensive with empirical science, at least by implication. That this is the result of a foreshortened view of the scope of human reason was pointed out at length by a colleague of mine, Dom Oliver Grosselin, in his paper on "The relation of physics to philosophy."[5]

This limited perspective leads to an interesting dilemma when the question of society is considered.[6] If the sciences of phenomena are the only source of certain knowledge, there can be no science of society. For society is a whole whose unity is one of purpose or objective, and this objective is the common good of its members.[7] But values and ends are not investigated by empirical science. Thus, in this view they must be found by some other means than unaided natural reason; an appeal must be made to sacred theology, or to animal faith, vital force or whatnot, in order to explain the values and ends according to which society is constituted. But since practical experience teaches us not to expect universal agreement except through reason, society becomes a practical impossibility because of the lack of a discoverable common purpose. Society can be held together in that case only by invoking persuasion, force, or a fuzzy kind of good will.

Fortunately the case is not quite so hopeless as this. The intention presiding over what follows is to suggest that a wider view of science makes it possible for natural human reason to

---

* Paper delivered at the Pennsylvania Conference of College Teachers, October 25, 1947.

[1] A. Einstein, quoted in the December 1946 issue of the *Reader's Digest Discussion Patterns*.

[2] A. Einstein, from *The world as I see it*, reprinted in "My faith" series, *Am. Weekly Mag.* (Oct. 13, 1946).

[3] A. N. Whitehead, *Science and the modern world.* (Cambridge, 1926); Penguin ed., p. 14.

[4] A. H. Compton, quoted in the December 1946 issue of the *Reader's Digest Discussion Patterns*.

[5] O. A. Grosselin, *Am. J. Physics* 9, 285 (1941).

[6] It is not my intention to impute the following consequences to the men quoted above, with the possible exception of Einstein as he has expressed himself in "Science and religion," *Science, philosophy and religion, I* (New York, 1941), p. 209.

[7] See St. Thomas Aquinas, *Contra impugnantes Dei cultum*, 3, c.: " . . . societas nihil aliud esse videatur, quam adunatio hominum ad unum aliquid communiter agendum." This is a unity of *order:* " . . . hoc totum, quod est civilis multitudo, vel domestica familia, *habet solam unitatem ordinis*, secundum quam non est aliquid simpliciter unum." *In X Libros Ethicorum Aristotelis*, I, lect. 1, No. 5.

discover enough about the ends of human life to provide a rational basis for the structure of human society, and that for this discovery empirical science alone does not suffice. Whether or not this natural investigation will disclose the complete fulfillment or end of the human person without recourse to sacred theology is a question we do not touch here.[8] Let us see how far reason alone will take us.

An analysis of how *order* is related to our reason provides an excellent means of avoiding confusion, and of bringing our problem into focus. Order may be defined as the distribution of things according to a principle or purpose which determines whether they come before or after one another, above or below, and so on.[9] A comparison of order with the human intellect or reason reveals four distinct kinds, as St. Thomas Aquinas reminds us at the beginning of his commentary on Aristotle's *Ethics*.[10]

First of all, there exists an order that reason does not create, but simply considers or knows; this is the order of the natural world, of the universe. There is another kind of order that reason, in the process of knowing, introduces into its own proper activity, as, for example, when it establishes order among its concepts, judgments and demonstrations so as to keep knowledge in conformity with reality, thus maintaining the relation called truth. The activity of reason sets up still another kind of order when it regulates the operations of the will or appetite. And, finally, reason establishes a fourth kind of order when it rearranges material things within the possibilities of their own proper natures, imposing an arbitrary form upon them.

Now, since the working of our reason is brought to its complete or highest stage by habit—and science is an intellectual habit—the diversity of sciences is a consequence of these diverse kinds of order which the intellect considers. The order of the universe, which reason considers but does not make, is the object of natural philosophy and metaphysics, and of course of the sciences of phenomena. The order which reason introduces into its own activity belongs to rational philosophy, that is, logic. The order of voluntary acts belongs to the consideration of moral philosophy, that is, ethics, including social and political ethics. Lastly, the order which reason causes to be imposed on external things pertains to the mechanical arts and the crafts, or art in the widest sense. The latter three kinds, it may be noted, are all in some way derivatives of the first, the order of the universe, in which all our knowledge originates.

Since our present interest is in the scientist and the common good, only the logical and ethical orders concern us directly, though the others must be kept in mind. Logic will indicate the interrelation and scope of the sciences, and ethics will tell us about the scientist as a complete human person—a rational, free and political animal by his very makeup.

A science is any body of correctly coordinated knowledge that reflects to the fullest possible extent some aspect of the ordered universe. One may say that it really gives us the universe in our knowing. (Einstein has aptly called it "the attempt at the posterior reconstruction of existence by the process of conceptualization."[11]) Each science is autonomous, and is the only competent authority in its own field. This field is delimited by the object which it studies and the method by which it proceeds. Each science, with the exception of metaphysics, begins with initial hypotheses or assumptions which it need not and even cannot demonstrate. The science then develops within these hypotheses according to the first principles of the operation of reason. This results in a hierarchical order among the sciences, so that the first principles of one science are guaranteed by a superior science, that is, one with a wider range and within the scope of which these principles fall.

Metaphysics, a valid science of the real, begins with no hypotheses, but only the first principles of knowing. Because its object is *being* itself, its conclusions are at once the most universal and the most central, since everything partakes of being. Thus one may compare the whole scheme of the sciences to a sort of pyramid, with metaphysics at the top.[12] The point to be insisted

[8] See M. J. Adler, "The future of democracy," *Proc. Am. Catholic Phil. Assoc.* **20**, 1 (1945).

[9] See St. Thomas Aquinas, *Summa Theologiae*, II–II, qu. 26, art. 1.

[10] *In X Libros Ethic. Arist.*, I, lect. 1, No. 1–3.

[11] A. Einstein, reference 6, p. 209.

[12] I realize the antagonism this comparison can arouse if not properly understood. It refers to the inverse variation

upon is that metaphysics, though the most universal, does not, as Descartes thought, absorb the other sciences; nor is metaphysics merely a synthesis of the particular sciences, as with Auguste Comte and the positivists.[13] *Each science remains autonomous* because of its peculiar principles and method.

The much misunderstood regulative function of the superior sciences over the inferior is not in any sense arbitrary or fascistic. Each science may develop indefinitely on its own level and never conflict with another. "Superior" and "inferior" may be here understood, according to the analogy of the pyramid, with reference to their distance from the apex. This position is determined by the nature of the science itself, as has been said. The regulative function merely refers to the fact that when a given science arrives at a *conclusion* which passes beyond its competence, that conclusion is in another field, and is to be judged by the principles of that field. There is no question of one science being a criterion of the *methods* of another. The proper method of a science is intelligible only within the limits of that science. The question is solely one of *conclusions*.

This, I am aware, is a recklessly and dangerously condensed account of an extremely complex subject. Perhaps an illustration or two will be reassuring.

Surely no one would send a set of surgical instruments to a chemistry laboratory and expect them to accomplish their purpose. But this is too obvious. Likewise no one would expect a mathematician to explain the nature of quantity, or its relation to the being of which it is a quantity. He merely accepts quantity as axiomatic and goes to work with it. Again, no pure mathematician is able by his science to arrive at conclusions concerning the nature of man or his ultimate ends, the notion of good, or of causality; much less of essences and natures of existing things. They are simply out of reach of his mathematical instruments. Conversely, it would not be true that two quantities equal to a third are equal to each other, unless it were first true that two beings identical with a third are identical with each other. Nor, to carry it a step further, would it be true that if you have two eggs and get two more you have four eggs, unless two plus two always *equal* four in mathematics, and *are* four in metaphysics. If the legitimate efforts of an empirical science yield conclusions that contradict those of mathematics, the latter science has the right to object as being the more certain.[14] But both sides had better work the problem over again.

Theoretical physics, so magnificently developed in recent times, occupies a position both unique and precarious in the scheme of the sciences. It has for its object the entire material universe, but under the aspect of change, of quantity, quality, relation, and the combinations of these which make up the other variable characteristics of material things. It is thus unique in its peculiar union of the two levels of mathematics and empirical physics. It is precarious because it has as its object all observable things, yet does not reach the level of metaphysics, and consequently cannot tell us about being and its properties—truth, goodness, beauty—nor about essence and nature. Accordingly, it is inclined to attempt an explanation of these matters within its own limited sphere, anywhere between abstract quantity and sheer flux. This is because science is always found in a human subject, and the human intellect always has to grasp *being* before it can grasp anything else.[15] From then on, however, much as the scientist may abstract from being, its ghost will always haunt him. If he has decided that the empirical sciences reveal all that can be known, he will try to fit being into his scheme. A glance at Professor Einstein's article on "Science and religion" will illustrate what can happen then. God, or the Absolute Being, simply *is* the universe, and you then

of the comprehension and extension of concepts, as treated in logic. A concept which is more comprehensive (that is, includes more characteristics of its object) applies or "extends" to a smaller number of objects. Thus the concepts of the natural sciences have greater comprehension and less extension than those of mathematics. In the latter case, matter and change are excluded, and quantitative concepts thus apply in the same way to all the natural sciences. When quantity is also abstracted from, we are left with the simplest (least comprehensive) concept: that of *being*. The extension here becomes universal, since the concept *being* applies analogically to all that does or can exist. Here we are on the level of metaphysical science, and it is because of the universal validity of its concepts, judgments and conclusions that it is the "supreme" science, the apex of the pyramid. See J. Maritain, *The degrees of knowledge* (Sheed & Ward, 1938), part I, ch. 1.

[13] For Comte, of course, the *term* metaphysics has a quite different meaning, and refers to one of the pre-scientific stages of information. The question here is of the science which unifies and regulates the others, and it is in this sense that the term metaphysics is here used, regardless of the differences in the terminology in which Comte gives account of the synthesis of the sciences.

[14] More certain because free from the variable elements introduced by the matter and change which natural science includes, but from which mathematics abstracts.

[15] See the splendid treatment of this by G. B. Phelan, "Verum sequitur esse rerum," *Mediaeval Studies*, 1 (1939), p. 11. The article is in English, despite its Latin title.

spell it with a capital. He is the God of Spinoza. A transcendent, personal God is a silly, anthropomorphic product of imagination which we should have outgrown long ago.[16] Has the conclusion gone beyond the premisses again?

The difference in method between the philosophical and the empirical sciences will have another important result which we shall need shortly. Briefly, it is this: The philosophical sciences are primarily deductive, and deal with the intelligible (that is, *trans*-empirical) characteristics of beings. In considering the essences and natures of things they arrive at conclusions causally. The laws which they state are thus *necessary* laws; they are intrinsically and inevitably true. The laws of empirical science, on the other hand, may or may not imply necessity. At any rate, they are treated as hypothetical, and always subject to revision. They are statements of uniformity, succession, relation, usually capable of mathematical formulation and without insistence on causality. This is because the empirical sciences remain within the sphere of quantity, quality, relation, and so on, and are not intended to reach beyond quantity to essence and nature.[17]

So much for science. Let us now pass to the scientist, the human person equipped with the intellectual habit of science. What is his relation to society?

So far we have seen that the empirical sciences do not cover the whole of human knowledge. We have now to notice that human knowledge, however complete, is only a part of human activity. Apart from putting order into his knowing, he has the problem of putting order into his doing. Here the whole man is involved, and he must widen his horizon to include philosophical knowledge to regulate his doing. To pursue ethical problems on exclusively empirical grounds will result in disaster, because the specifically human characteristics of reason and freedom are beyond the reach of empirical analysis. Values and ends are involved in accounting for human behavior, and since these are consequent upon essences and natures, we must call upon a science that considers these. In this case it will be philosophy, for we have limited ourselves in this discussion to rational inquiry. Here our first question must be, What is a man?

Reference has already been made to Aristotle's definition of man as a rational, free, and political animal.[18] This means that he can think, can recognize ends or objectives, and can choose and apply means toward the attainment of them. Because of his intelligence and freedom he is a personal being, an autonomous whole. And because he cannot accomplish the full limit of his development as a human person without the assistance of others, he is a social and political being. These characteristics are written into man's nature; that is, they are consequences of his very makeup. Reason, freedom and incommunicable personality are all capable of philosophical demonstration, without appeal to theology. But since space restricts us, and since there are grounds for assuming that my readers agree that we are at least potentially thinking and free persons—how else should we explain our reading scientific journals?—we shall now consider the notion of society.

Man cannot come into existence or remain there very long without the cooperation of other men. He makes his appearance as a dependent animal, and within some sort of association—normally the family. He is part of a whole, and the satisfaction of his wants is circumscribed by the needs of that whole. In technical language, his individual good is subordinated to a larger and common good, the order of the family. After a relatively long time his specifically human qualities appear: thinking and rational order in his voluntary acts. He is dependent on those who have already made progress in these things for the beginning of his education or, better, his natural fulfillment.

A glance at the history of civilization will assure us that no single family can supply all the means for the complete realization of human possibilities. For this a great number of men united by a common objective is required. This unity of objective is what constitutes *any* society.[19] A perfect human society would have as its end the provision of the greatest possible

[16] A. Einstein, reference 6, p. 211.
[17] See J. Maritain, reference 12, pp. 41–44.
[18] Aristotle, *Politics*, Book I, ch. 1–2.
[19] Reference 18, I, 1; also St. Thomas' commentary on this section of the *Politics*, and reference 7.

good for all its members. The common good that constitutes society thus implies the establishment of conditions under which equal opportunity is afforded each member for living the "good life," as it has been called. The practical structure that effects these conditions is the political order, and the right to establish this or that particular kind of order resides in the members of society as free and intelligent persons.

The common good *is* superior to the individual good, as the good of the whole is superior to the good of a part. This, however, must be understood carefully in the case of human good. Society is a whole, and we are its parts.[20] But society is not an organic whole; its unity is solely one of purpose, and each of its parts is an autonomous person and a complete unit. Professor Maritain has called society a whole composed of wholes.[21] Thus if the common good is not at the same time the good of the persons and does not, so to speak, flow back on them in greater measure than they could achieve independently, *it is not the common good.* It is the opposite, for it checks instead of developing human persons. In so doing it defeats the purpose of society and of government and automatically becomes a tyranny.

The importance of reaching agreement on the nature of man should now be apparent. Like all other beings, he seeks the complete realization of his possibilities. He differs from all other material beings in having the ability to recognize these possibilities and to choose means that bring about their realization. Intelligence and freedom are his alone, and these attributes are what make him a man.[22] The acquisition of truth and virtue are thus his highest natural objectives. But these cannot be completely accomplished in material terms; that is, his "end" is not adequately described in material terms, although it necessarily includes them.

We have agreed not to ask whether man has also a supernatural end. He could not discover this by natural means, and we are now concerned with society on the plane of nature. Such a supernatural end would add to and elevate man's natural end, but would change it. Civil society founded on an analysis of human nature and ends will take us in the right direction. If it is not far enough, God Himself must show us the rest of the way, and our only caution to the political scientist is not to block that way.

We are now prepared to draw some conclusions about the scientist and the common good which determines society. First of all, let us consider his contribution as an empirical scientist, a specialist; let us say a physicist. In this role he provides man with materials and instruments that can be used with tremendous effect to further or to destroy the common good. He can give us control over nature; he can give us an atomic bomb. But since his science can tell us nothing about good or evil, he cannot tell us what to do with this control, or when to use the atomic bomb. His professional contribution is *amoral*, and as such carries with it neither obligation nor responsibility.[23]

However, it is not as a scientist but as a human person that the scientist (or anyone else) enters society. This fact has profound significance in a democratic society, where all the members have a share in determining the structure and policy of the government. Here it is important for all the members to understand and to develop the whole of their human nature, and for some of the members to understand the philosophical science of government particularly well. Plato wanted the philosophers to be kings.[24] Now we no longer need kings, but we do need legislators who know human values and are skilled in political ethics in proportion to their responsibility in governing men, in guiding them to their fulfillment.

Consider the case of legislation as an illustration. The distinction has been made between the law in empirical science and the natural law. The one is statistical and hypothetical, the other essential and necessary. If the laws that regulate human society are formulated on an empirical basis, the specifically human element is disregarded, and it is a matter of chance whether or not they conflict with the intrinsic laws of human nature. Yet it should be clear

[20] See St. Thomas Aquinas, *Summa Theologiae*, I–II, qu. 90, art. 2, c.

[21] J. Maritain, *Scholasticism and politics* (Macmillan, 1940), p. 72.

[22] *Summa Theologiae*, I–II, qu. 31, art. 7, c.

[23] Of course, as Dushman said at the conference with so much force, the scientist does have the obligation of pointing out the consequences of his discoveries and inventions.

[24] *Republic*, Book VII, 520*b*.

from the argument of this paper that civil law presupposes natural law, and is the further determination of the latter according to such variable circumstances as time, place, character of the people and so on.[25] To make laws on an empirical basis is to treat *persons* as simply units in an arithmetic sum. This procedure has no logical escape from materialistic communism, fascism or some other violent form of totalitarianism which knows nothing of personal rights. The common good is hypostasized as The State, and sometimes even deified. Men become mere humanity through an error in human logic.

* * *

There are other things that the scientist does for society in addition to offering the fruits of his work, making sure his scientific perspective is not restricted, and endeavoring to have life more abundantly. Some of these are often unconscious. Who would venture to estimate the effect on others of his habit of clear and accurate thinking? Or of his unselfish and prolonged labors to produce the things and the knowledge that have made our civilization what it is?

Those of us who are engaged in education can contribute *pro rata* to the realization of an even more important benefit to society. We can do our best about restoring an integrated human education which develops the whole man first of all, and the specialist afterward, as Chancellor Hutchins and Professor Adler have been urging us to do for so long. For, if we forget what it means to be a man, who will argue that we deserve the survival of a human civilization?

[25] *Summa Theologiae*, I–II, qu. 95, art. 1, c.

Production and Measurement of Air Bubbles in Water*

C. H. TINDAL AND D. C. WHITMARSH
Ordnance Research Laboratory, The Pennsylvania State College, State College, Pennsylvania

IN investigations of the transmission of energy through water, it is often desired to determine the effects of air bubbles in the water. In some cases it is necessary to have a volume distribution of bubbles; in others it is better to produce the bubbles in a line and allow them to form a vertical sheet as they rise to the surface. Usually bubbles of a prescribed size are wanted. Although in practice the sizes will vary somewhat the distribution of the sizes may be determined.

Production of Bubbles

Production of bubbles in water by blowing air through orifices has many inherent difficulties. Sufficiently precise control of the air pressure to yield bubbles of a reproducible size is difficult to obtain, and even if it is, the degree of wetting of orifice edges may actually be the factor controlling the bubble size.

It is well known that heating aerated water will cause the evolution of air bubbles. For the study of energy transmissions, it is necessary that the temperature of the water shall remain practically constant. A convenient method of producing bubbles is to immerse a wire of Nichrome or other alloy of high resistivity in the water and heat the wire electrically with either direct or alternating current. The power dissipated in the wire may be determined by ammeter and voltmeter readings. Figure 1 shows a dense curtain of bubbles produced in a large tank by a 17-in. helix of wire, heated by a current of 45 amp at 110 v. For a volume distribution of bubbles a grid or screen may replace the wire. It should be mentioned that more power is needed with such an arrangement. Often a zigzagged single wire can produce sufficient bubbles in a small volume.

The rate of bubble formation may be measured by collecting the bubbles and measuring the volume of air collected in a given time. Figure 2 shows a type of collector tube found convenient for collection of bubbles from a 2-cm length of

* This work was sponsored by the U. S. Navy Department under Contract NOrd 7958.

wire for 5-min intervals. An inverted burette, equipped with a funnel, will serve the same purpose.

The rate of bubble production is a function of the degree of aeration of the water and of the power input to the wire. Saturated aeration, being a function of temperature and pressure only, is desirable, and may be insured by the simple apparatus shown in Fig. 3. This aeration tube is made of an 8-in. length of 1-in. glass tubing, with a piece of 45-mesh wire gauze in the middle. Water is run in at the top and out at the bottom; air is passed through it in the opposite direction. The air inlet tube is longer than the water outlet, so that water will not drain into the air inlet when the air supply is shut off. Air at a pressure of 10 lb/in.^{2} is provided by a regulated, motor-driven, paint spray pump. Three such aeration tubes in series are sufficient to guarantee saturated aeration, as long as room temperature and the barometric pressure remain reasonably constant.

The size of the bubbles is independent of the power dissipated. However, the rate at which bubbles are produced increases with input power. This is shown by Fig. 4. It will be noticed that the rate of bubble liberation is linear up to a certain power input above which the rate increases rapidly. It is of interest to note that this rapid increase is accompanied by an audible "singing" sound.

Measurement of Bubbles

To develop a technic for the measurement of the size of very small gas bubbles in water, various standard methods, such as direct photography and projection onto a grid, were tried and found to have severe shortcomings. Problems of illumination and suitable camera positioning were encountered in direct photography. Projection was useful in determining the order of magnitude of bubble diameter, but bubbles rising through a liquid move so rapidly that any accurate determination of the size by this method is impossible. Attempts to photograph the screen upon which the bubbles were projected ended in failure because of inadequate light.

A method of direct photography by projection was finally adopted. The technic is to project the bubble images upon a screen, immediately in front of which is placed a camera equipped with shutter but no lens.

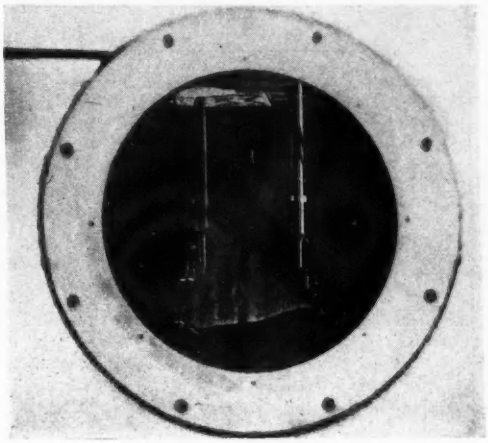

FIG. 1. Bubbles from a hot wire.

The first set of equipment assembled for trying this method consisted of a Bausch and Lomb slide projector, a Lucite tank in which the bubbles were produced, a Speed Graphic camera, and film in film holders. The tank was made of $\frac{1}{8}$-in. sheet Lucite; its dimensions were about $4\times5\times1$ in., so that it would fit the opening in front of the condenser lens of the projector when the slide holder was removed. The projector lens focused the bubble images on a white screen.

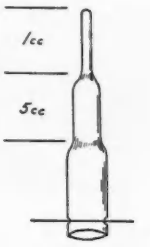

FIG. 2. Collector tube.

FIG. 3. Aeration tube.

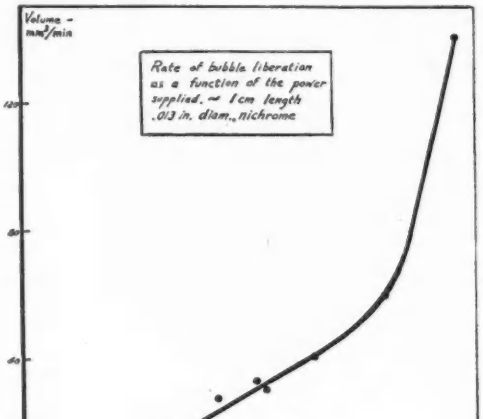

FIG. 4. Bubble liberation rate from a hot wire.

The back of the camera was placed in the middle of the field against the screen. By watching the screen the bubbles which were desired could be photographed when they came into the field of the camera. Exposures as short as 0.001 sec were used and produced excellent bubble images.

For calibration purposes a transparent scale was placed in the tank so that it could be photographed simultaneously with the bubbles. This system proved quite successful for determining the size of bubbles as they were produced by a hot wire or by electrolysis with wire electrodes. Figure 5 is a typical photograph of bubbles made with this arrangement. This shows bubbles forming on a hot wire at very low power. The large bubbles on the wire were made by the union of many small bubbles. After convection currents become established, small bubbles leave the wire, as seen at the top of the picture. Later it became desirable to use larger bubble producers in a tank deeper than could be contained within the projector. A tank having over-all dimensions of about $8\times10\times12$ in. was therefore constructed of $\frac{1}{4}$-in. Lucite sheet. The objective lens tube was removed from the projector and mounted in a special frame in front of the tank. The position was so chosen that the bubbles were photographed as they rose beyond the bubble-producing mechanism. For a light source, a single No. 2 Photoflood lamp in a metal housing was used. The housing prevented stray light from reaching and fogging the film; the tank was enclosed in a cardboard box with holes for the light beam for the same reason. The results were as good and the equipment was much more flexible than the slide projector.

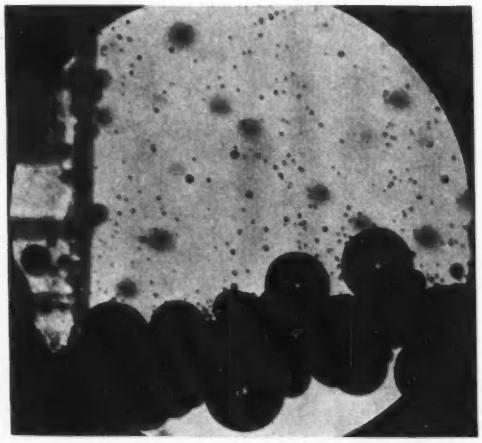

FIG. 5. Typical bubble photograph made by the projection method; the vertical scale at the left is divided in millimeters.

Some trouble was experienced because the bubbles did not always rise vertically from the bubble-producing device, and from a large bubble generator the bubbles might be formed over a fairly large area so that all the bubbles in a cross section of the bubble stream would not be in focus on the screen simultaneously. Since the bubbles that were to be measured were rather small, a system for catching them was developed. A piece of clear $\frac{1}{8}$-in. Lucite plate was held in the water over the bubble generator or moved through the water above the generator until a sufficient number of bubbles had collected on the plate. The bubble generator was then turned off and the Lucite plate maneuvered into position in the field of the objective lens so that the bubbles on the plate were focused on the screen. This permitted more time for photographing the bubbles and produced negatives that were easy to measure. A transparent scale again was photographed simultaneously with the bubbles for calibration purposes. Figure 6 is a bubble photograph taken of bubbles caught on a Lucite

plate. These bubbles were made by forcing air through an unglazed porcelain filter.

When the bubbles were produced in a tank which was not suitable for projection, the Lucite plate could be maneuvered into the small projection tank, which was submerged in the larger tank in which the bubbles were produced. The small projection cell could then be placed in the projector and the bubbles could be recorded. As long as the collecting plate was not removed from water, the bubbles were retained on the plate.

It was found that with this lens system in the small room where this work was done (about 15 ft long) it was possible to get a magnification of about 20 from the bubble to the negative. If the negative, after it was developed, were reprojected in the slide projector, an additional magnification of 20 could be obtained which gave a magnification of 400 diameters from the original bubble.

All of this work was done in a darkened room, of course. It was found, however, that a photographic room was not necessary, for with the film placed in the beam of the projector considerable external light could be tolerated without having the film fogged noticeably at the exposure speeds used.

Although this method has been used to photograph bubbles produced in water, the bubbles being either stationary or in motion, the application is not limited to this use. Anything capable of being projected can be photographed and measured in this way. This method may be used with microprojectors in biology or for photographing anything in water or other transparent liquids. It can also be used for recording solids which are capable of being projected, as in experiments on crystal growth.

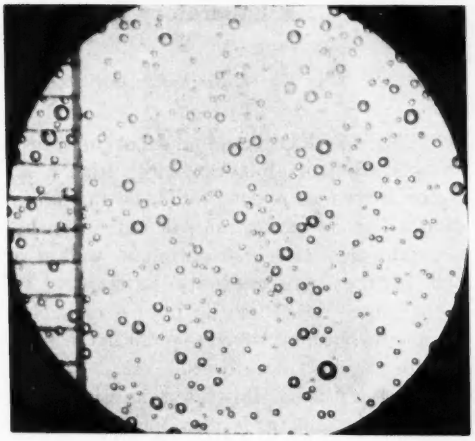

FIG. 6. Typical photograph of bubbles caught on Lucite plate; the scale is divided in millimeters.

Progress in physical theory has from time to time come to a standstill during periods in which the theory has become increasingly complex and incapable of accounting for the observed phenomena, and which have eventually been resolved by the emergence of a fundamentally new idea. Examples of such new ideas are the electromagnetic theory of light, the older quantum theory, the principle of relativity and the new quantum mechanics. The present is again a time of failure of physical theory. The impasse has lasted about ten years, but the required new idea is still lacking.—M. H. L. PRICE, *Nature* **160**, 627 (1947).

A Laboratory Experiment on Alpha-Particle Scattering

SANBORN C. BROWN
Massachusetts Institute of Technology, Cambridge 39, Massachusetts

RUTHERFORD'S original experiment on the scattering of alpha-particles, using a scintillation screen as a detector, was difficult and tedious. The newer technics of detection, however, put this classic experiment within the reach of the undergraduate laboratory. The present paper describes the details of such an experiment that has proved popular in our undergraduate physics laboratories.

The apparatus is illustrated in Fig. 1. The alpha-particle detector is a proportional counter. The entire experiment is carried out under a bell jar through which helium is flowing. This procedure has several advantages. The proportional counting range for the detector occurs at voltages below 1000 v for helium at atmospheric pressure. The pressure is made very slightly above atmospheric; there is no trouble with leaks, since helium escapes through any joint that is not completely tight. The slight excess pressure in the system is maintained by the depth of the liquid in a bottle through which the helium is allowed to bubble. Any light oil such as Nujol is found satisfactory for this purpose. The empty bottle in the line prevents oil from being sucked back into the apparatus should the helium flow stop and a change in temperature reduce the pressure below atmospheric. The range of alpha-particles in helium at atmospheric pressure is long enough to make the apparatus of convenient size.

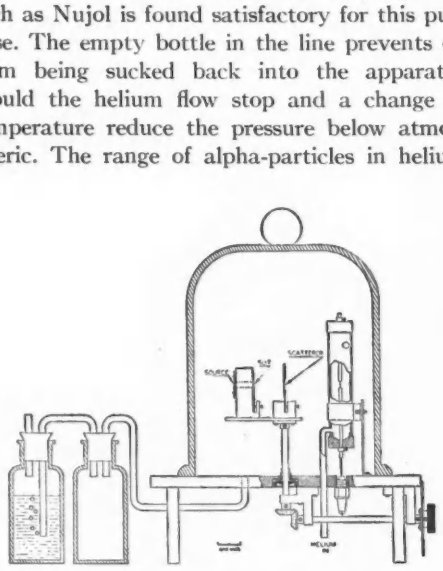

FIG. 1. Alpha-particle scattering apparatus.

The source of alpha-rays used is about 0.25 millicurie of polonium plated on a small piece of nickel foil.[1] The alpha-rays are collimated by a slit and scattered by a thin gold foil. This scatterer is of the type of gold leaf ordinarily used by sign painters.

The scattered radiation enters the counter through a slit in the cathode wall $\frac{1}{16}$ in. wide and $\frac{1}{4}$ in. high. The sensitive volume of the counter is restricted to a central section where the anode wire is $\frac{1}{4}$ in. long and 0.008 in. in diameter. This fine central wire is supported by $\frac{1}{8}$-in. brass rods. This counter construction restricts the sensitive volume, thereby minimizing the counting background, but maintains enough uniformity of electric field within the counter to give good proportional-counter characteristics.

The anode wire of the counter is connected to the preamplifier circuit shown in Fig. 2 through a spark-plug insulator screwed through the base plate of the apparatus. The preamplifier is mounted directly below the spark plug. The output of this circuit leads to an ordinary scale-of-64 counting circuit and message register.

The source holder and scatterer are mounted on a rotating shaft which is fed through the base plate by a Wilson seal,[2] of the type used by the Air Forces for rheostat and switch shafts on pressurized radar equipment. The angle through which the source and scatterer is rotated can be read on a calibrated dial on the front of the base plate.

The purpose of the experiment is to show that the net counting rate N is proportional to the scattering angle θ, according to the equation

$$N = K \csc^4 \tfrac{1}{2}\theta.$$

A typical student run taken with this apparatus is shown in Fig. 3. The number of

[1] Obtained from the Canadian Radium and Uranium Corp., 630 Fifth Ave., New York, N. Y.
[2] Made by the Galvin Manufacturing Company, Chicago, Ill.

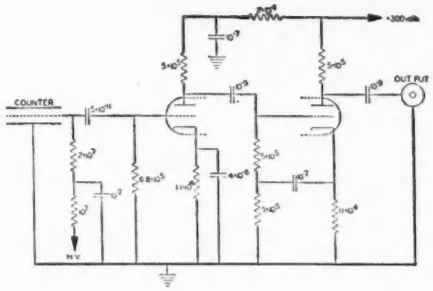

FIG. 2. Proportional-counter preamplifier circuit. The values of the resistances are in ohms and those of the capacitances are in farads.

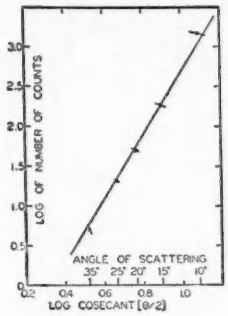

FIG. 3. Typical experimental data. The line through the points has a slope of 4.

counts recorded in a fixed time interval is plotted against both the actual scattering angle and logarithm of the cosecant of half the angle. The background counting rate was 1 count/min. The probable errors are indicated on the points to illustrate the accuracy of results obtained in a 3-hr laboratory period. The straight line has the theoretical slope of 4.

A Method for Measuring Surface Tension

A. A. EL-KARIM
Farouk I University, Alexandria, Egypt

A METHOD is described for measuring the coefficient of surface tension and the density of a sample of liquid by the same apparatus. The surface tension can be measured without knowing the density. The method is to adjust the pressure in the apparatus until the meniscus in a capillary tube is at a distance equal to one-third of the radius of the capillary above or below the surface of the liquid in an adjoining glass cell.

Apparatus

The arrangement of the tensiometer for a liquid that wets glass is shown in Fig. 1. One end of capillary tube C, the bore of which should not exceed 1 mm in diameter,[1] is connected rigidly to a glass cell A_1 with a plane wall and a volume of about 7 cm³; the other end is connected by means of rubber tubing to a glass manometer U and to a bottle B_1. The pressure of air in B_1 can be increased slowly by allowing water to enter it from another vessel.

[1] A. Ferguson and S. J. Kennedy, *Proc. Physical Soc.* **44**, 511 (1932).

The liquid whose surface tension is to be measured is put in A_1 and rises in the capillary to a height h above the liquid level in the cell.

When the pressure is slowly increased the meniscus will be forced down. It is then carefully adjusted until the meniscus is at a distance $\frac{1}{3}r$ below the plane surface of the liquid in the glass cell. This position may be judged with very considerable accuracy by watching through a vertical

TABLE I. Surface tension of various liquids.

Liquid	Radius r of tube (cm)	Density ρ of manometer liquid (g/cm³)	Temperature θ (deg C)	Coefficient of surface tension T_θ (dy/cm)
Water	0.021	0.99913	15	71.66
	.036			71.9
	.045			72.35
Benzene	.021	.99823	20	30.2
	.036			29.93
	.045			29.81
Acetone	.021	.99862	18	23.54
	.036			23.39
	.045			23.19
Mercury	.051	1.021	18	547.1
	.045			546.8
	.036			546.9

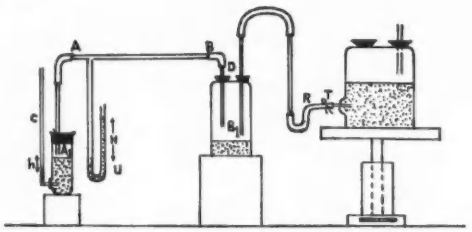

FIG. 1. Apparatus for measuring surface tension of liquids that wet glass.

reading microscope carrying a vernier graduated to 0.01 mm. The mean radius r of the capillary tube is found from the weight of mercury required to fill it.

For a liquid that does not wet glass, the capillary tube is connected as shown in Fig. 2, and the meniscus is adjusted to a level $\frac{1}{3}r$ above that of the liquid in A_1.

Theoretical Considerations

For a *concave* meniscus, whether it is above or below the plane surface of the liquid in the glass cell, the equation for the coefficient of surface tension T is[2]

$$T=\tfrac{1}{2}rg[\rho H\mp(h\mp\tfrac{1}{3}r)\sigma],$$

where ρ and σ are the densities of the manometer liquid and the liquid under test, respectively, and H and h are the heights shown in Fig. 1. If, then, the pressure is changed until h becomes equal to $\frac{1}{3}r$ and the meniscus is below the plane surface, the equation becomes

$$T=\tfrac{1}{2}rg\rho H.$$

For a *convex* meniscus, as in the case of mercury, whether the meniscus is below or above the plane surface in the glass cell, the equation is

$$T=\tfrac{1}{2}\frac{rg}{\cos\alpha}[-\rho H\mp(h\pm\tfrac{1}{3}r)\sigma],$$

which becomes

$$T=-rg\rho H/2\cos\alpha$$

when the meniscus is $\frac{1}{3}r$ above the plane surface.

Thus a knowledge of the density of the liquid under test is not necessary for finding T. However, if H, and consequently h, be varied, a graph

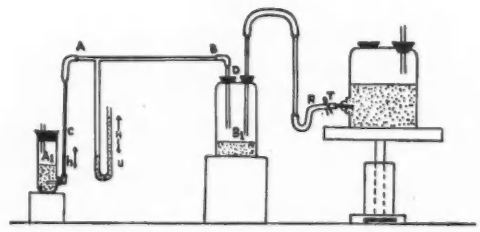

FIG. 2. Modification of apparatus for liquids that do not wet glass.

[2] A. Ferguson, *Proc. Physical Soc.* **36**, 37 (1923).

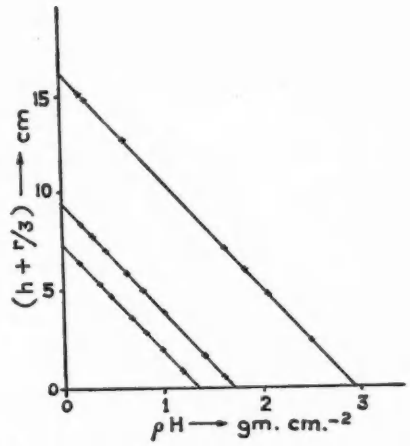

FIG. 3. Variation of $h+\frac{1}{3}r$ with ρH when the meniscus is above, for benzene.

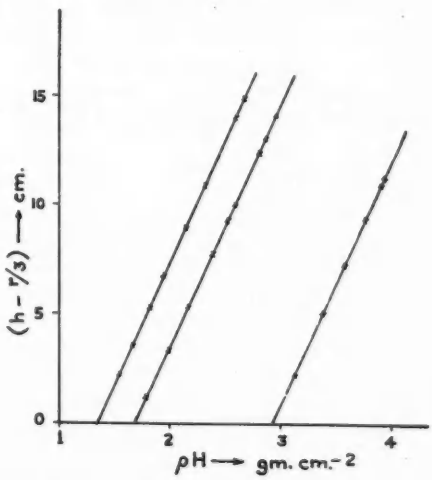

FIG. 4. Variation of $h-\frac{1}{3}r$ with ρH when the meniscus is below, for benzene.

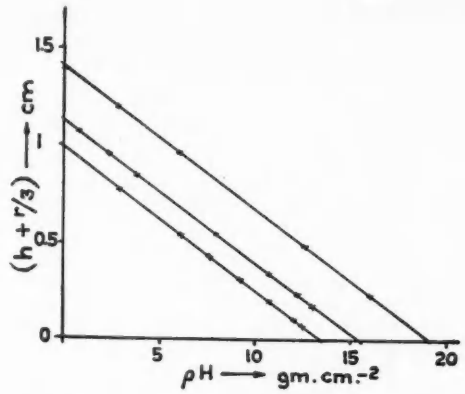

FIG. 5. Variation of $h+\frac{1}{3}r$ with ρH when the meniscus is below, for mercury.

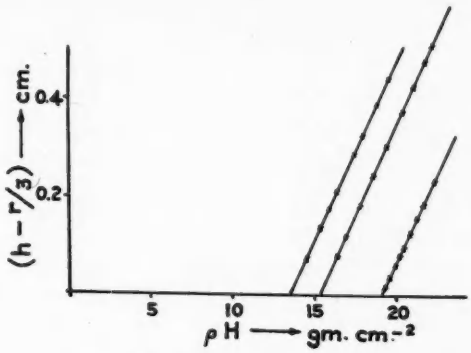

FIG. 6. Variation of $h-\frac{1}{3}r$ with ρH when the meniscus is above, for mercury.

of ρH as a function of $(h\mp\frac{1}{3}r)$ gives a straight line, the slope of which gives $1/\sigma$ and the intercept on the ρH-axis gives $2T/gr$.

Experimental Results

Experiments have been carried out on water, benzene, acetone and mercury. The substances were of first-grade research quality and were free from the common impurities. The work was done at room temperature, which remained sensibly constant during the half-hour required to take any one set of observations. Three tubes were used, having radii of 0.045 cm, 0.036 cm and 0.021 cm, respectively. The results are shown in Table I.

Figures 3, 4, 5 and 6 show the graphs obtained for benzene and mercury with each of the three tubes. The group of lines for each substance should of course be parallel because the slope of the graphs is the same.

I wish to express my thanks to Dr. M. A. El-Sherbini for his valuable help and advice during the progress of this work.

New Members of the Association

The following persons have been made *members* or *junior members* (*J*) of the American Association of Physics Teachers since the publication of the preceding list [*Am. J. Physics* **16**, 252 (1948)].

Marie, Sister Gabrielle, Dunbarton College of Holy Cross, Washington 8, D. C.
Manning, Walter H., Jr., Pennsylvania State College, State College, Pa.
Miasnick, Joseph D. (*J*), 40 Wayland St., Dorchester 25, Mass.
Miller, David C., Montana State University, Missoula, Mont.
Millman, George H. (*J*), Box 558, State College, Pa.
Murphy, James A. (*J*), 210 S. Sparks St., State College, Pa.
Murtaugh, Rev. Walter A., Providence College, Providence, R. I.
Musser, Glen L., 132½ E. Prospect Ave., State College, Pa.
Nader, John S., 6271 Orchard Lane, Cincinnati 13, Ohio.
Newlin, Wm. H. (*J*), 840 Montgomery Ave., Bryn Mawr, Pa.
Nielsen, William F. (*J*), 112 N. Duck St., Stillwater, Okla.
Oleck, Robert F., 143 Burhans Ave., Yonkers, N. Y.
Palubinskas, Felix S., Iowa State College, Ames, Iowa.
Pappo, H. Alexander (*J*), 3519 Fourth St., Detroit, Mich.
Perry, David De Wolf (*J*), 25 E. Upsal St., Philadelphia 19, Pa.
Peter, Bro. Simon, Saint Mary's College, Saint Mary's College, Calif.
Prugh, Edwin K., University of Dayton, Dayton 9, Ohio.
Pulskamp, C. A., 12 Inman Ave., Kentfield, Calif.
Ransom, R. H., 6201 Third Ave., Ro emount, Apt. 7, Mo real, Canada.
Read, Albert J. (*J*), Rensselaer Polytechnic Institute, Troy, N. Y.
Reiber, Daniel G., 231 S. Atherton St., State College, Pa.
Rix, Herbert D., 1020 W. Beaver Ave., State College, Pa.
Rollosson, George W., 516 E. Fourth St., Crowley, La.
Schwartz, Milton (*J*), 5016 N. Franklin St., Philadelphia 20, Pa.
Scibor-Marchocki, Romuald I. (*J*), 120 Florence Ave., Highland Park 3, Mich.
Seren, Leo, University of Idaho, Moscow, Idaho.
Sheehan, John A. (*J*), 141 Aldrich St., Roslindale 31, Mass.
Tarris, Demetrius, 60 Owego St., Simpson, Pa.
Thompson, Kingsley P., 1780 Delmont Ave., East Cleveland 12, Ohio.
Walker, John P., Jr. (*J*), Box 518, State College, Pa.
Weichardt, Heinz H., 49 Williamson Ave., Bloomfield, N. J.
Wittman, E. Frances, 3530 Fulton St., San Francisco 18, Calif.
Wolfram, Orland A., Alma White College, Zarephath, N. J.
Woodbridge, Caspar L. (*J*), Pennsylvania State College, State College, Pa.

Discovering the Torsion Pendulum Expression in the Freshman Laboratory

SEVILLE CHAPMAN
Stanford University, Stanford, California

DURING the postwar reorganization of the laboratory in the first-year Engineering Physics course at Stanford University, approximately half of the experiments have been replaced by more substantial ones. Although in most cases new equipment has been built for new experiments, the old equipment for the torsion pendulum experiment (Fig. 1) has been retained, but an entirely new approach to the experiment has been introduced. In unsigned ballots, the students voted the experiment in its new form the most valuable one in the course in mechanics.

Until 1946, our laboratory instruction sheets gave the calculus derivation of the formula for the period P of the torsion pendulum as a function of the moment of inertia I of the plate, the length l of the wire, the radius r of the wire, and the shear modulus η of the wire, namely,

$$P^2 = 8\pi Il/\eta r^4. \qquad (1)$$

The experiment consisted in having the students "verify" the formula by varying the several parameters to see if the results agreed with Eq. (1). For the freshmen who had not yet had calculus, the derivation was mysterious, and the experiment had little meaning.

In the new laboratory manual,[1] Eq. (1) is not given, but the students are asked to *discover* it by experiment and by reason. Fortunately the formula appears explicitly in few books that first-year students are likely to study. During the lecture the students are shown a torsion pendulum, and how it works. They are asked to study a dimensional argument in the laboratory manual, to review the section on graphs, and to bring graph paper to the laboratory.

As stated in the laboratory manual, some of the factors that possibly might be thought to influence the period of the torsion pendulum are:

Length of the torsion wire.
Radius (or diameter) of the torsion wire.
Density of the torsion wire.
Shear modulus of the torsion wire.
Young's modulus for the torsion wire.
Mass of the plate.
Density of the plate.
Thickness of the plate.
Moment of inertia of the plate.
Moment of inertia of the rings on the plate.
Weight hanging on the torsion wire.
Acceleration due to gravity.
Atmospheric pressure.
Room temperature.
Color of the paint on the apparatus.
Amplitude of vibration.
Numerical constants.

The present experiment consists in investigating several of these factors.

The students begin by measuring the masses of the plate and of the two rings, the inner and outer diameters of the two rings, the outer diameter of the plate, and the diameters of the wires. They then carry out five subsidiary experiments.

(1) The torsion pendulum is set up, and the effect of amplitude of vibration on the period is investigated. The period is found to be independent of amplitude.

(2) It is found that the period depends on the length of the wire. A graph of period as a function of length is a curve, but (period)^2 against length is a straight line, showing that (period)^2 is proportional to length.

(3) The effect of a weight hanging on the torsion wire is investigated. When a weight hangs by a torsionless string from a hook in the center of the bottom of the plate the period is not affected, but if a ring of the same weight is placed on the plate, the period is altered.

(4) Two rings of calculable moment of inertia are used, and four measurements of the period are made: (i) when no rings are on the plate; (ii) when the smaller ring is on the plate; (iii) when the larger ring is on the plate; and (iv) when both rings are on the plate. The students plot a graph of period as a function of calculated moment of inertia of the rings added to the plate. The result, of course, is a curve, but various powers of the period may be plotted against various powers of the moment of inertia until a straight line is obtained.

[1] S. Chapman, *Laboratory manual, engineering physics* (National Press, Millbrae, California).

(5) The effect of the radius of the wire is investigated. Students may plot (period)^2 against various powers of radius to get a straight line, or they may plot (period)^2 against radius on log-log paper and measure the slope to get the proper exponent for the radius.

The radius of the wires is difficult to measure accurately, and only the best students get results that are definitive. For this reason all students are required to make a dimensional argument involving the radius of the wire, the modulus of elasticity of the wire, and the acceleration due to gravity, all of which may be suspected of being significant. Since the students know already that (period)^2 is proportional to the length of the wire and to the moment of inertia of the plate plus rings, the dimensional argument shows that (period)^2 is proportional to (radius)$^{-4}$ (modulus of elasticity)$^{-1}$ (acceleration due to gravity)^0.

The dimensions of Young's modulus and the shear modulus are the same. Therefore an optional experiment is available for those students who wish to make an experimental determination by which to decide which modulus is appropriate. If a suitable comparison is made between the periods for brass and for steel wires of equal length with a given number of rings on the plate, the proper modulus can be found. If E is the modulus of elasticity, then

$$\frac{P^2_{\text{steel}}}{P^2_{\text{brass}}}=\frac{E_{\text{brass}}r^4_{\text{brass}}}{E_{\text{steel}}r^4_{\text{steel}}}. \qquad (2)$$

From experimental data the ratio $E_{\text{brass}}/E_{\text{steel}}$ can be found. The value of this ratio is different for Young's modulus and the shear modulus. If the experimental work is done carefully, a clear indication is obtained as to the proper modulus. (Some readers may raise the logical objection to this experiment that to determine the values of the moduli, the laboratory instructors must have worked the experiment backward, but the students seem unaware of this objection.)

The constant of proportionality in Eq. (1), namely, 8π, is determined from the slope of one of the straight-line graphs; most students get between 23 and 27. Except for the optional experiment, the laboratory work takes the average students about 2.5 hr.

For the students the whole experiment furnishes a good example of the application of the scientific method to determine new knowledge. Most of them enjoy it very much, although there are a few who dislike it intensely, and ask why they must waste so much time to find a formula they could obtain much more easily if we would tell them what handbook to look in.

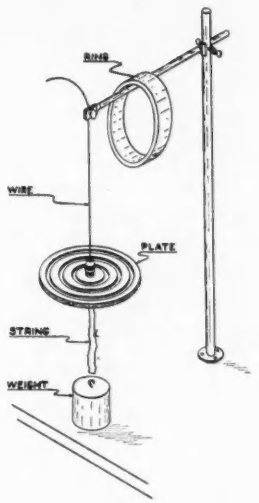

FIG. 1. The torsion pendulum experiment (reproduced from reference 1).

In order that an experiment of this type be effective, the final expression must not be readily available to the students. This requirement precludes applying the method of this experiment to many important laboratory problems. The members of the instructing staff believe that the experiment has worked out very well. We have thought of an experiment to determine the force constant of a helical spring in terms of its geometry and shear modulus, and of experiments involving viscous liquids—either capillary flow, or spheres falling in accordance with Stokes' law. We feel that these experiments are related somewhat too closely to the torsion pendulum experiment to warrant their inclusion, but we are trying to invent some experiments of this type for other parts of the course.

A man who has committed a mistake and doesn't correct it is committing another mistake.—CONFUCIUS.

Measurement of the Mutual Forces between a Magnet and a Wire Carrying a Direct Current

ALVA TURNER
Hunter College, New York 21, New York

IT is well known that if a bar magnet is placed in the magnetic field of a direct current in a wire, each pole of the magnet experiences a force and that for each force acting on the magnet there is an equal and opposite force acting on the wire. It may therefore be concluded that the resultant force acting on the magnet because of the current in the wire is accompanied by an equal and opposite resultant force acting on the wire because of the magnet. It is an easy matter to demonstrate this qualitatively, but until recently the quantitative measurement of the two forces has not been so simple. With modern magnets, made of alloys such as Alnico V, it is now possible to measure simultaneously the magnitudes of the mutual forces. The apparatus here described may be used for lecture demonstration or for a student experiment.

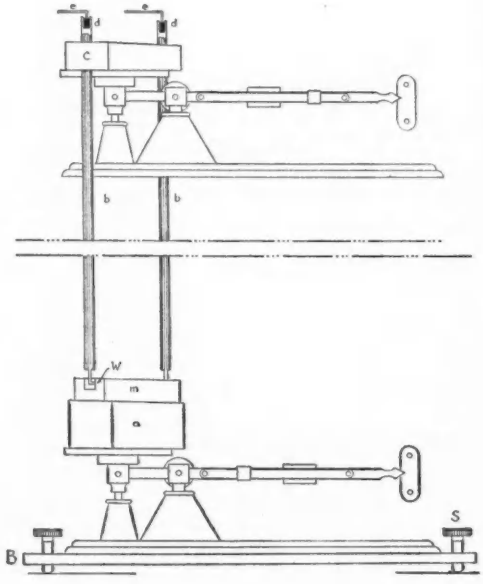

FIG. 1. Two-balance apparatus for measuring the mutual forces between a magnet and a wire carrying a direct current.

Apparatus

The apparatus is shown in Fig. 1. Two heavy-duty, triple-beam balances are held in position, one vertically above the other, by a wooden frame, not shown in the figure. The upper balance supports the wire *W* and the lower one supports the magnet *m*. The magnet (Alnico V, No. 502–6) is 15.4 cm long, 2.5 cm wide and 2.10 cm high, with an air gap between the pole faces 15.4 cm long, 0.95 cm wide, and 1.00 cm deep. The magnet is supported by a wooden block *a*, which is fastened to the pan of the lower balance. This balance is mounted on a wooden base *B*, provided with leveling screws *S*, so that the magnet can be adjusted to the desired position relative to the wire.

The wire *W* is 15.4 cm long and 2.14 mm in diameter and is suspended 45 cm below the top of the pan of the upper balance by means of two copper rods *b*, 6.4 mm in diameter, supported by the wooden spacer *C*, which is fastened to the pan with two screws. Two No. 16 copper wires *e*, connect mercury cups *d*, at the tops of the rods *b*, to binding posts which are fastened to an upright on the wooden frame. The mercury cups make it possible to adjust the upper balance without altering the current in the wire. The currents needed do not exceed 15 amp.

Procedure

Both balances are adjusted for equilibrium when there is no current in the wire. The lower balance is adjusted until the wire is in the optimum position in the air gap of the magnet. This optimum position may be different for forces of repulsion and of attraction. When a current is established in the wire it is necessary to readjust the balances for equilibrium. The necessary change in weights in the lower balance gives the magnitude of the resultant force acting on the magnet, and the necessary change in weights in the upper balance gives the magnitude of the resultant force acting on the wire.

To facilitate the actual taking of data it is found convenient to proceed as follows. Five 5-g weights are placed on the pan of the lower balance, the wire is put slightly above the middle of the air gap, and both balances are brought to equilibrium. The direction of the current is made such as to produce attraction, and the magnitude of the current is adjusted to its maximum value; the beam weights are then used to restore the balances to equilibrium. A 5-g weight is then transferred from the lower balance to the upper balance, and the current is adjusted until one of the two balances is restored to equilibrium. The other balance is then adjusted by moving the rider, which gives measurable changes in weight of less than 1 g. (This requires a change in weight of 0.1 g or less in each case.) This procedure is repeated, a 5-g weight being transferred each time from the pan of the lower balance to the pan of the upper one.

If the direction of the current is such as to produce repulsion, the five 5-g weights are placed on the pan of the upper balance and transferred from the upper to the lower.

TABLE I. Mutual forces on a current-carrying wire and a magnet in its field.

I (amp)	F_m (gwt)	F_m/I (gwt/amp)	F_w (gwt)	F_w/I (gwt/amp)
11.65	25.50	2.19	25.40	2.18
9.43	20.50	2.17	20.40	2.16
7.02	15.37	2.19	15.40	2.19
4.80	10.48	2.18	10.40	2.17
		av. 2.183		av. 2.175

Results

The data in Table I were obtained for forces of attraction. Equally accurate results are obtained for forces of repulsion. In this table, F_m and F_w are the forces acting on the magnet and the wire, respectively. The table shows that the mutual forces are equal and that the force per unit current is the same for all currents used.

Jerome Cardan, 1501–1576

R. B. LINDSAY
Brown University, Providence 12, Rhode Island

IN the fall of the year 1575 an old man of 74, spending his last years as a comparative stranger in that turbulent city on the banks of the Tiber, sat down to cast up his accounts with the world and to provide an apology for his existence. One might have supposed he had written enough, since he had contrived already to fill 7000 folio pages of Latin on almost every subject under the sun. But it was clear that death alone could keep his restless pen from the paper and his equally restless mind from speculation. The result was an autobiography—not, to be sure, the greatest, but one of the most curious examples of self-analysis ever written by a great scholar.

In his prologue JEROME CARDAN says:

> This Book of My Life I am undertaking to write after the example of Marcus Aurelius Antoninus the Philosopher, acclaimed the wisest and best of men, knowing well that no accomplishment of mortal man is perfect, much less safe from calumny; yet aware that none of all the ends which man may attain is more worthy than recognition of the truth.

Posterity has had a hard time making up its mind whether the writer of these words was really an honest and zealous seeker after the truth or a hypocritical rascal who sought future fame by scribbling whatever nonsense came into his head and only by accident, as it were, occasionally "rang the bell." No one has ever doubted, however, that JEROME CARDAN was a colorful personality and a figure worthy of the attention of physical scientists. His autobiography is not so full of exciting adventures as the more famous work of his artistic contemporary, BENVENUTO CELLINI, but it makes equally fascinating reading.

Who can fail to be interested in a man who, though known primarily as a successful practicing physician, published in 1545 a work in mathematics called by E. T. BELL "the sum and crown of all algebra up to his time!" This was the same man who practiced astrology to the extent of believing seriously that he would not live to see 45 and hence in early life wasted his substance in riotous living; later he wrote books discussing learnedly why his prediction had been wrong. Called by the great Italian renaissance jurist, ANDREA ALCIATI, "the man of discoveries" and by JULIUS CAESAR SCALIGER, a bitter critic

FIG. 1. Jerome Cardan, 1501–1576.

of some of CARDAN's writings, "a man of most profound, most favored and incomparable genius," and admitted to be the greatest popularizer of science of his day, he yet believed firmly that he was accompanied by a familiar demon or guardian angel which took care he would suffer no harm in the daily vicissitudes of life. Forever insisting that he was a teller of the truth, he managed nevertheless to get off his chest some amazing tales.

Let us make a brief excursion into the autobiography of CARDAN before we examine his views on physical science. He tells us that he was born in Pavia in the Duchy of Milan on September 24, 1501. He came into the world without benefit of clergy, and there is evidence indeed that he was decidedly not wanted. His father was a learned man, a juris consult by profession and forever advising everyone he met to "praise the Lord, who is the source of all virtue." CARDAN asserts solemnly that from the aspect of the stars he might easily have been a monster. As it was, he was a puny enough infant who caught nearly every childhood ailment there was to catch; perhaps the greatest miracle in his life was that he ever managed to survive at all. His home life while a boy cannot have been very happy. He says that from the age of seven he was forced to accompany his father as a servant, carrying law papers, and being kicked around generally.

His education was informal and spasmodic, being conducted largely by his parents. Not until he reached 19 was he allowed to attend the academy, which later became the University of Pavia. Here he was entered to study law, but soon gave this up in favor of medicine, not without considerable paternal opposition. It must be remembered that in those days the study of medicine involved the consideration of practically all the phenomena of nature. CARDAN's curiosity about the world around him had evidently been aroused very early. It is fair to say it never really left him for the remaining 50 years of his life, though it led him into some strange predicaments and odd speculations.

By the time he was 21, the young JEROME, so named by his father after the saint who was supposed to have saved his life during an attack of fever at the age of 8, had reached the point where he was able to give lessons in Euclid and undertake public disputations. In 1524 he transferred to Padua and after two rejections, possibly caused by the questionable legitimacy of his birth but quite as likely by his rather wild habits, including an inveterate love of the dice box, he was finally granted the degree of Doctor of Medicine in 1526. The year previous had witnessed the almost incredible incident of CARDAN's holding the rectorship of the University. At any rate, he claims this honor in several of his works, though no record of it has ever been found in the archives of the University.

After the attainment of his degree in medicine, CARDAN started practice in the village of Sacco, about ten miles from Padua. He would have preferred to reside in Milan, but at this time the big city was afflicted with a succession of plagues and was therefore a decidedly unhealthy place for a physician. It does not appear that CARDAN took his practice too seriously. If one may be-

lieve his own account, his stay in Sacco was a happy, rather indolent time. After all, he did not believe he was going to live long, and was naturally disinclined to take seriously the troubles of this world. While there were dice to cast, good food to eat, plenty of amusing folk willing to eat it at one's expense and good music to listen to, one could at least ignore some of the more obvious aspects of life's wretchedness. We must remember that all his life CARDAN was plagued with an interesting assortment of bodily afflictions which he catalogs in his autobiography with a thoroughness approaching affection. No wonder he occasionally contemplated suicide!

CARDAN's stay in Sacco was, however, notable in one respect: there he first began that career of scribbling which led ultimately to the enormous production of written material to which reference has already been made. His first attempts, as might have been expected, were in the field of medicine.

But it was not until he had married and moved to Milan about 1531 that he turned to writing in earnest. He was deprived by the Milanese medical association of the legal right to practice medicine, presumably because of his illegitimate birth, though his admitted tendency to rub people the wrong way by emphasizing the disagreeable rather than the agreeable in conversation may have accounted for much of the coolness with which he was received by the medical profession in Milan. He and his family would probably have been reduced to beggary had it not been for the influence of a friend who secured CARDAN's appointment to a private endowment for the giving of lectures in geometry, arithmetic and astronomy. He immediately began to write on these subjects. At the same time, to eke out his slender earnings as a lecturer, he gave medical treatment to those who would have him despite his nonprofessional status, and achieved some noteworthy cures. The latter naturally did not endear him to his medical brethren. Nor did his first printed book, published in Venice in 1536, *On the Bad Practice of Medicine in Common Use*, make them more eager to welcome him with open arms. It was only the beginning of a long series of irritations to his contemporaries, which make it rather surprising that CARDAN did not come to a violent end.

In his autobiography CARDAN explains that he took up writing to perpetuate his name, while at the same time he rather sardonically remarks it was really futile to hope that anything he wrote would ever endure for more than a few years and that all his pains would go for naught. Actually this was mere rationalization. CARDAN suffered from *cacoethes scribendi* and wrote because he could not help it. Besides mathematics, of which we shall have more to say presently, he wrote on astronomy, including selected genitures, physics, morals, medicine, music, theology and divination. He discoursed with equal enthusiasm on the uses of adversity and in praise of NERO; on poisons and on seven-month parturition; on metoposcopy, or the art of reading character from the lines in the forehead, and on imaginary numbers; on the teeth and on games of chance, in which last he was able to speak from wide personal experience. In addition to the 131 printed works which have survived, CARDAN says he burned 170 in manuscript. Small wonder that he actually found it necessary to write a book *about* his books!

By the time he was 40 CARDAN finally received a medical license in Milan and thereafter achieved considerable success in his practice and his lecturing, both in Milan and Pavia. Indeed, as he himself says, "It was when I ought to have died that I really began to live." Books began to flow freely from his pen, and celebrated people consulted him about their ills. The correspondent and friend of the famous anatomist, VESALIUS, the ardent correspondent though not exactly the friend of the mathematician, TARTAGLIA, CARDAN began at length to take his rightful place among the intellectually great. Attracting the attention of foreign dignitaries, he was offered a high post with a tempting salary at the court of Denmark, but declined because he was afraid the climate of Copenhagen would be too rigorous. Besides the Danes were Protestants! Curiously enough he did accept an invitation to go to Edinburgh to attend on the ARCHBISHOP OF ST. ANDREWS, brother of JAMES HAMILTON, the Regent of Scotland in the infancy of MARY, later Queen of Scots. The ARCHBISHOP had a bad case of asthma which threatened to carry him off.

During his stay in Edinburgh of nearly three months in the summer of 1552, CARDAN managed to relieve the patient's discomfort and persuade him to a more healthful way of living. At any rate the Milanese physician could boast that the ARCHBISHOP never died of asthma; he went out by a route all too common in Scotland in those days—the scaffold.

It would be pleasant to be able to report that after his early trials CARDAN was rewarded by a life of comfort and distinction. Unfortunately, his troubles were not over when he achieved professional success. Plagued incessantly by ill health and irritated by the effects on his oversensitive temperament of the envy of contemporaries, he received his worst blow when his eldest son, who had contracted an unfortunate marriage, was convicted of wife murder and executed for the crime. This was in 1560. There is reason to believe that the father never really recovered from the terrible shock. He grew steadily more superstitious, and life became one bad omen or portent after another. Yet he kept on writing to the end. In a dream a voice told him he could secure forgetfulness of his son's fate by putting in his mouth an emerald that he commonly wore on a chain around his neck. And it worked! But as a psychiatric measure it was not too practical, and the old man reverted again and again in his later writings to the terrible tragedy that had befallen his house. He died in Rome on September 20, 1576, convinced to the last that he had been a "teller of the truth, an upright man, and indebted for his powers to a divine spirit."

In the contemplation of the intellectual activities of CARDAN we shall here limit ourselves to his mathematical and physical contributions and shall not discuss his achievements in medicine. As a matter of fact, it does not appear that he made any discovery in medical science of lasting significance, though he certainly was a successful practitioner. On the other hand, his work in mathematics was of real consequence for the subsequent development of physical science, and this despite the fact that the famous method for solving the cubic equation now known by his name was actually due to TARTAGLIA OF BRESCIA and published by CARDAN without the discoverer's consent, though with full acknowledgment of his identity. The story of the great controversy connected with this is of sufficient interest to stand retelling.

CARDAN's greatest work is admitted to be his *Liber Artis Magnae Sive de Regulis Algebraicis* (1545), the "Book of the Great Art or the Rules of Algebra." Six years previously he had published a rather elaborate work on arithmetic, containing a fair amount of what might be called algebra from the geometric point of view. Up to the year 1500 the quadratic equation was the algebraic equation of highest degree for which anything like a general solution had been discovered. Even here only positive roots were admitted to have meaning. Of algebraic symbolism as we know it there was none. The cubic equation was tantalizing, and it was not until 1505 that a native of Bologna named SCIPIO FERREUS discovered a rule for handling the special case of the cubic equation $x^3+bx=c$, b and c being positive integers. As has just been pointed out, the equation was not expressed in modern symbolism. Rather the statement was expressed in words which when translated into English would say: the cube of a thing plus b things equals the number c. FERREUS divulged his method to a pupil, a Venetian named FLORIDO, who kept the matter secret after the custom of those days and enjoyed challenging his contemporaries to algebraic duels in the solution of such equations, which he usually won with ease. One day FLORIDO had the bad luck to challenge NICCOLO FONTANA, better known as TARTAGLIA (because he stuttered), a native of Brescia domiciled in Venice. TARTAGLIA was a self-taught mathematician who had mulled over these algebraic questions and had found out a thing or two about them. He succeeded not only in unraveling the cubic that FLORIDO presented but also managed to work out some cases that FLORIDO had never learned. The result was rather sad for the Venetian, but made TARTAGLIA's fame secure.

CARDAN, then engaged on the preparation of his mathematical treatise and learning of the famous competition between FLORIDO and TARTAGLIA, addressed the latter through a bookseller friend and requested the favor of his rules for solving the various forms of the cubic. He wished to publish them in his book, making all

due acknowledgment to the discoverer. On the other hand, if TARTAGLIA preferred, he would keep them secret. The stutterer sent back word that when he got ready he would publish his discoveries himself and not in another man's book. He included a few other nasty remarks tending to cast doubt on CARDAN's standing as a mathematician. This naturally irritated MESSER JEROME, who sent back a red-hot answer which TARTAGLIA thought so much of that he included it in his own published reminiscences, along with his version of the whole controversy which was just then beginning. In his reply CARDAN offered to wager 100 ducats that TARTAGLIA could not answer certain questions. He also threw in a gratuitous criticism (as a matter of fact well founded) of a point in a recently published work of TARTAGLIA's on ballistics. This might have ended the matter had not CARDAN had the cleverness to include among his castigations a reference to the fact he had presented a copy of TARTAGLIA's work on artillery to the MARQUIS DEL GUASTO, one of CARDAN's wealthy patrons in Milan. TARTAGLIA was so wrapped up in ballistics that he could not resist the temptation to forward for presentation to the MARQUIS some new instruments for use in artillery firing. This gave CARDAN an excuse for writing again in more temperate vein, begging TARTAGLIA to come to Milan, since the MARQUIS wished to see him.

The meeting of the two scholars finally took place in Milan on March 25, 1539. The conversation must have been interesting. CARDAN pleaded for the divulging of the magic rules and swore "by the sacred gospel and on the faith of a gentleman" that he would never publish or otherwise divulge them, that he would moreover set them down in a cipher so that after his death no one could possibly understand them. The attitude toward the dissemination of scientific discovery inherent in this famous colloquy of two of the outstanding intellectuals of the time deserves more consideration than we can give to it here. Suffice it to say that TARTAGLIA was persuaded by CARDAN's oath and finally gave him a little poem in the Venetian dialect in which the rules are very cryptically concealed. He then left for home quite convinced, apparently, that he had done the wrong thing and that he was about to be double-crossed. He was right; when the "Great Art" appeared in 1545 it contained the famous rules with full acknowledgment to TARTAGLIA but no mention of the agreement thus unceremoniously broken.

There is no doubt that CARDAN was guilty of deceit. But there is also no doubt that the advance of mathematics was hastened by a dozen years or more, since TARTAGLIA did not bring out his own work on algebra until 1556, and even then it was not equal to the thoroughgoing treatise of CARDAN. When one thinks of the sort of crimes that were commonplace during his time it is difficult to be harsh with CARDAN in this matter. He at any rate possessed the modern spirit, which insists that scientific discoveries are the common property of all those who are willing to take the trouble to understand them, as long as their authorship is freely acknowledged.

CARDAN's eminence in mathematics strongly suggests that a look at his achievements in the realm of physics would be worth while. He wrote voluminously on this subject. His longest single work is the 21 books of the *De Subtilitate*, commonly translated as "On Subtlety," first published in 1550. Of these the first seven books are devoted to what we should today call physics, whereas the remaining 14 run the gamut from plants and animals through man to God and the universe, with some respects paid along the way to dreams, objects of art and demons. CARDAN relates in the eighteenth book of *De Subtilitate* how he saw in a dream a work containing precisely 21 parts on varied subjects with a little geometry in the middle and written in a delightful style with great subtlety of reasoning and not without a certain pleasing obscurity. To write it seemed beyond the power of mortal man. But he finally did it and filled up 315 large folio pages of Latinity in the process. Examination of the work indicates that the prescription of obscurity was fairly faithfully realized. CARDAN notes with considerable satisfaction that he dreamed about the work during the whole period of three years devoted to its composition and only stopped when the book was finally printed in Nuremberg!

The *De Subtilitate* and the associated 17 books of the *De Rerum Varietate*, "Concerning a Variety of Things," form almost literally a

medieval *Science for the Citizen:* what the cultivated renaissance gentleman ought to know about the world. To the 20th century scientist it appears to be a queer medley of fact and superstition, of moderately accurate description and absurd explanation. Thus, for example, when CARDAN considers the siphon, he describes its observed behavior quite as accurately as we expect to find it in a modern book on mechanics, but the explanation involves much talk about heavy water tending to fall and light water to rise. The pressure of the atmosphere receives no comment.

Though to be fashionable CARDAN disagreed with ARISTOTLE on a number of points, it seems clear that his views on physics were still largely Aristotelian in character. He claimed great credit for reducing the number of fundamental elements from the conventional four—earth, water, air and fire—to three by discarding fire from the list. The reason given was that fire is a form of motion and motion cannot be an element, since it is not a substance. The verbalization here is distinctly a part of the Aristotelian heritage.

In astronomy the writings of COPERNICUS evidently had no influence on CARDAN though he must have been aware of the existence of the *De Revolutionibus* when he composed his *De Subtilitate*. He believed that the earth is round, is fixed at the center of the universe, and does not move as a whole. All this, he says, is proved by mathematics. Apparently CARDAN was not the first nor the last to manifest sublime confidence in the "proofs" of mathematics. He did not go so far as to claim that the earth is absolutely immovable, since he recognized the existence of earthquakes, whose origin he attributed to the explosive burning of combustible materials within the earth's crust.

Like most medieval philosophers, CARDAN had a passion for classification and subdivision. Thus, he argues that all prime matter (the stuff of which everything else is made) is divided into two classes, namely, celestial and terrestrial, and hence all phenomena exemplify these two fundamental varieties. The fundamental celestial property is the heat of the sun and stars, which is the origin of all terrestrial heat. This sounds moderately promising to a modern scientist, but when CARDAN then goes on to talk about the celestial heat as the world soul, we find ourselves wondering just what he means and wherein his speculations constitute any advance on ARISTOTLE.

Of course we can play with CARDAN the same pleasant game that has been played by historians of science on many other early philosophers: we can read into his language our own ideas of what he may have meant and thus attribute to him scientific views quite complimentary to his insight into the nature of things. When, for example, he locates the origin of fire in the conjunction of the celestial heat with a dry object and the origin of the natural heat of life in the conjunction of celestial heat with a moist object, then come to mind ARCHIMEDES and his burning glass on the one hand and the little plant growing in the moist earth under the warm sunlight on the other, and we think: how clever of CARDAN! But the trouble is that this sort of thing does not get anywhere; it predicts nothing and suggests no operations for testing the statements so glibly made.

Similar remarks apply in even more emphatic fashion to CARDAN's views on comets. He says they always cause winds and, through their influence on the air, bring about the death of physically weak and voluptuous persons as well as the aged and those afflicted with insomnia! We are entitled to imagine that he is, after all, only spoofing here, for he goes on to point out that the preceding characteristics are often found united in princes and this explains why they are apt to die suddenly when a comet appears. Moreover, comets usually produce droughts and are therefore indirectly responsible for pestilence and war. Perhaps we should not be too hard on CARDAN when we recall the recent attempts at the correlation of sunspot activity with economic and social phenomena.

In the fifth book of *De Subtilitate*, in the midst of a general description of different sorts of substances found in the earth, CARDAN introduces a few remarks on the curious tendency of rubbed amber to attract light pieces of straw and compares this property with the somewhat similar attraction of a magnet for iron. He gives no sources and does not say whether he had tried the experiments himself. There is little to indicate in his writings that he had arrived at anything like the conception of experimentation which his distinguished fellow countryman GALILEO was to develop half a century later. At any rate, the facts he presents are more or less cor-

rect and possibly the most extensive, accurate statement of the phenomena of electrification by friction up to his time. As usual, he feels impelled to give reasons for the behavior of the amber; so he says, "the attraction of the amber is without doubt like that of little gourds for hot things and due to the warm fat contained within it." Reasoning by analogy unchecked by experiment was accepted without question by the average medieval citizen, and CARDAN's contemporaries probably experienced no discomfort in swallowing such statements. One wonders whether the average modern reader of popular physics is any more critical.

Even a brief examination of CARDAN's career from the standpoint of scientific methodology prompts a fundamental question: How did it happen that he was able to get his mathematics straight enough to work out what were for his time extremely intricate problems in algebra and geometry and yet fail to apply the same clarity and efficiency of reasoning to problems involving natural phenomena?

The contrast strikes one most forcibly in his *Opus Novum de Proportionibus*, published in 1568. Freely translated, the contents are said to relate to the "proportions of numbers, motions, weights, sounds and other measurable things not only established by geometric reasoning but also illustrated by ingenious demonstrations involving all kinds of experiments and observations of natural phenomena." This might almost pass for an introduction to theoretical physics! And, indeed, as long as he sticks to the relations of numbers everything proceeds clearly and plausibly enough, albeit mostly via EUCLID's *Elements*. But when suddenly in proposition 23 he begins to discuss motion, the fabric disintegrates into the same old rubbish immortalized by ARISTOTLE, and all genuine application of mathematics disappears. Thus, CARDAN proceeds to reclassify motions and attempts to improve on ARISTOTLE by introducing *three* categories: ***voluntary motion***, always circular, as in the case of the heavenly bodies; ***natural motion***, always in a straight line, as in the case of a freely falling body; and ***violent motion***, which can take place in a curve, as in the case of a projectile. The explanation of the difference is very simple; voluntary motion is "in loco," natural motion is "ad locum," and violent motion is "ex loco!" Small wonder that CARDAN never discovered the simple rule for the behavior of a body on an inclined plane.

On second thought, the mystery of why a clever mathematician like CARDAN was not able to apply mathematics successfully to the development of mechanics and hence to the founding of theoretical physics is merely a special case of the greater problem of why the human race needed some 1800 years to evolve physical concepts to which the methods of pure mathematics could be successfully and efficiently applied. After all, applied mathematics demands a different attitude of mind from pure mathematics. Up in heaven, if that is where he is, CARDAN probably holds the view that it was inscrutable fate which determined that it was not he but GALILEO who was destined to be the founder of theoretical physics. We should not be too hard on him; he entertained his contemporaries, and he enjoyed himself. We may as well take as sincere the avowal in his autobiography:

> One thing alone provides satisfaction to me: to understand and grasp the meaning of all the wonders of the world would be more precious to me than the everlasting dominion of the universe; this I swear by all that is holy.

The fact that by our standards he did not succeed in this aim gives us no right to a supercilious sneer at his expense. We too, have still far to go to reach the goal that he laid down.

I remember, one day, saying how uphill the work was, and he [Thomson] answered, "Yes, that is why there is so much credit in doing anything."—RAYLEIGH, *The life of J. J. Thomson.*

Reproductions of Prints, Drawings and Paintings of Interest in the History of Physics

39. Hollar's Engraving Commemorating the Founding of the Royal Society

E. C. WATSON
California Institute of Technology, Pasadena 4, California

THE first history of the Royal Society of London was written by THOMAS SPRAT (1635–1713), the English divine who was elected a Fellow in 1663 and who later became Bishop of Rochester. It was first published in 1667, and a few copies contained a frontispiece engraved by the versatile etcher, WENCESLAUS HOLLAR (1607–1677), to commemorate the founding of the Society.

This famous engraving shows CHARLES II being crowned with a wreath, together with FRANCIS BACON (1461–1626) and Lord BROUNCKER (1620?–1684), in an open hall cluttered with scientific books and instruments. It was designed by the diarist, JOHN EVELYN (1620–1706), who was one of the promotors of the scheme for the Royal Society and on the king's charter of 1662 was nominated a member of the directing council. EVELYN, who was Secretary of the Society in 1672, was an intimate friend of Lord BROUNCKER, the first President of the Society, and the two often discussed scientific questions with CHARLES II. BROUNCKER was re-elected President each year until 1677.

Erratum. In my reply to Professor Gaehr's note on "Equations for Straight Lines" I made the statement: "If an odd number be made, the first or last of the series must be thrown out, which would be hard to justify" [*Am. J. Physics* **16**, 189 (1948)]. Professor Gaehr has pointed out in a letter to me that the middle observation is the one that should be omitted as this makes the differences to be averaged larger and therefore more reliable.—C. D. COOKSEY.

NOTES AND DISCUSSION

Franklin's Discovery of the Electron

ROBERT A. MILLIKAN
California Institute of Technology, Pasadena 4, California

IN celebrating the anniversary of the discovery of the electron, it is a mistake to concentrate attention solely on J. J. Thomson's paper of 1897. I do not in any way underestimate the importance of that paper in bringing about general acceptance of the electron theory of matter, through the evidence presented in it for the existence of a charged particle of *mass* about 1/1000 the mass of the hydrogen atom. However, to this particle Thomson gave the name "corpuscle," rather than electron, doubtless because the word "electron" had already been assigned by G. Johnstone Stoney to the hypothetical *atom of electricity*, that is, *of electric charge, whether that charge is positive or negative.* The existence of this atom of charge, or of the atomic constitution of electricity, was asserted with great clarity by Benjamin Franklin because of experiments begun in 1747, so that 1947 is both the 200th anniversary of Benjamin Franklin's discovery of the existence of an atom of electric charge and the 50th anniversary of J. J. Thomson's proof of the existence of a charged particle whose mass, as he then phrased it, was about 1/1000 that of the hydrogen atom.

These two conceptions have unfortunately been confused in much modern writing, though Thomson was careful to use the word "electron" in its original sense as an atom of charge, not of mass, and to call cathode rays a stream of "corpuscles." The most authoritative writers have continued to use the term "electron" to mean an *element of charge*, as G. Johnstone Stoney first defined it in publications in 1874, 1881 and in 1891.[1]

How unambiguously Franklin conceived his theory of electricity and described the most important properties of the electrical atom is shown by the following quotation from his letter to Peter Collinson in his book published in 1774.[2]

1. The electrical matter consists of particles extremely subtile, since it can permeate common matter, even the densest metals, with such ease and freedom as not to receive any perceptible resistance.

2. If any one should doubt whether the electrical matter passes through the substance of bodies, or only over and along their surfaces, a shock from an electrified large glass jar, taken through his own body, will probably convince him.

3. Electrical matter differs from common matter in this, that the parts of the latter mutually attract, those of the former mutually repel each other. Hence the appearing divergency in a stream of electrified effluvia.

4. But though the particles of electrical matter do repel each other, they are strongly attracted by all other matter.

5. From these three things, the extreme subtilty of the electrical matter, the mutual repulsion of its parts, and the strong attraction between them and other matter, arise[s] this effect, that, when a quantity of electrical matter is applied to a mass of common matter, of any bigness or length, within our observation (which hath not already got its quantity) it is immediately and equally diffused through the whole.

6. Thus common matter is a kind of spunge to the electrical fluid. And as a spunge would receive no water if the parts of water were not smaller than the pores of the spunge; and even then but slowly, if there were not a mutual attraction between those parts and the parts of the spunge; and would still imbibe it faster, if the mutual attraction among the parts of the water did not impede, some force being required to separate them; and fastest, if, instead of attraction, there were a mutual repulsion among those parts, which would act in conjunction with the attraction of the spunge. So is the case between the electrical and common matter.

7. But in common matter there is (generally) as much of the electrical as it will contain within its substance. If more is added, it lies without upon the surface, and forms what we call an electrical atmosphere, and then the body is said to be electrified.

I think the foregoing quotation clearly establishes the right of Benjamin Franklin to be considered the discoverer of the atom of electricity. *The world has recently and properly celebrated the year 1947 as both the 200th anniversary of Franklin's discovery of the electron and the 50th anniversary of J. J. Thomson's unambiguous establishment of the electron theory of matter.*

[1] See also R. A. Millikan, *Electrons (+ and −), protons, photons, neutrons, mesotrons and cosmic rays* (Univ. of Chicago Press, 1947), pp. 332–333.

[2] *Experiments and observations on electricity, made at Philadelphia in America, by Benjamin Franklin, L.L.D. and F.R.S. Member of the Royal Academy of Sciences at Paris, of the Royal Society at Göttingen, and of the Batavian Society in Holland, and President of the Philosophical Society at Philadelphia. To which are added, letters and papers on philosophical subjects. The whole corrected, methodized, improved, and now collected into one volume, and illustrated with copper plates.* (London, ed. 5, 1774), p. 54. The quotation may also be found in I. B. Cohen's edition (Harvard Univ. Press, 1941), p. 13.

A Convenient Viscosity Apparatus

G. P. BREWINGTON
Lawrence Institute of Technology, Detroit 3, Michigan

THE apparatus described herein has been found to be quite satisfactory for sophomore students and is sufficiently sensitive to indicate a change in the viscosity of

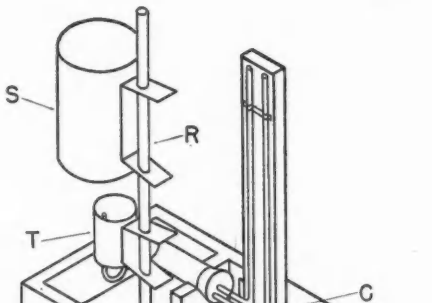

FIG. 1. Viscosity apparatus.

water as a function of temperature during the course of the experiment. The coefficient of viscosity is calculated by determining the factors appearing in the Hagen-Poiseuille law.[1]

The apparatus is designed to measure, by means of two manometers, the pressure differences between the ends of a capillary tube through which a steady stream of water is maintained. It differs from those described previously in that the complete equipment is mounted in a watertight tray (Fig. 1) which serves not only as a sink but a storage box. The upright metal rod *R* screws into a nut that is brazed or welded to a flat metal plate mounted on a small wooden platform, which in turn is fixed by screws to the bottom of the box. The constant-level tank *T* and the storage tank *S* are a soup can and a fruit-juice can. A small copper tube is soldered into the base of the storage can and is so pinched that the constant level tank functions continuously. The metal parts are painted with a good grade of black enamel. A capillary tube *C*, 1 mm in diameter and somewhat less than 6 in. long, gives good results with a hydrostatic head of 8 or 10 in. of water. A capillary tube of diameter 1.8 mm is unsatisfactory.

The holes in rubber stoppers may easily be enlarged to take the capillary tubes with an ordinary metal drill having sharp flutes along the sides. The drill selected should be somewhat larger than the capillary tube. A small V cut should then be made in the large end of the stopper to indicate that it has been altered.

The wooden strip carrying the manometer is removed by loosening one wing nut. The manometer scale is a strip of millimeter cross-section paper with every tenth line ruled black; this scale is cemented to the wooden upright by a plastic cement[2] known as "Acryloid B-7." This cement is clear, highly waterproof and, unlike shellac or varnish, does not seem to get brittle, crack or discolor. It should not be applied on surfaces painted with certain linseed-oil paints, for the chlorinated hydrocarbon solvent is an excellent paint remover. The interior of the box should be painted with a waterproof asphalt paint. Both the upright metal rod and the manometer are removable, and the base then serves as a storage tray for the complete apparatus. A great many construction details of this equipment were suggested by Mr. Michael George.

[1] Dodge and Thompson, *Fluid mechanics* (McGraw-Hill), p. 171; Hausmann and Slack, *Physics* (Van Nostrand, ed. 2), p. 212.
[2] Obtained from Rohm and Haas.

Mackay's Model of the Climbing Monkey

W. W. SLEATOR
University of Michigan, Ann Arbor, Michigan

WHEN I saw the brief note by Mackay[1] about the yo-yo as a working model of the climbing monkey I undertook to construct one. Professor Dennison and others contributed ideas, and after a few trials a successful

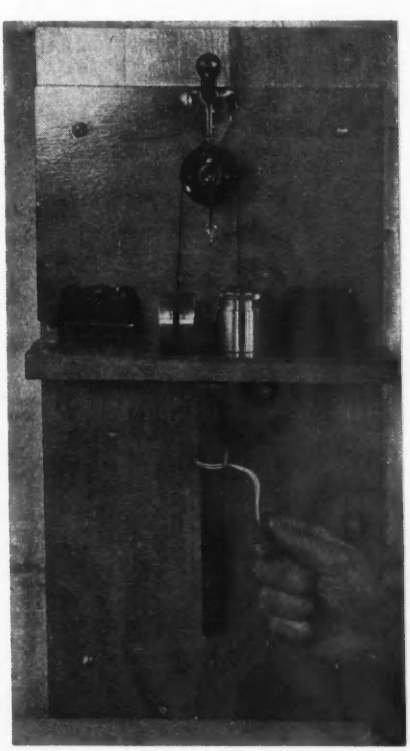

FIG. 1. All set! The spring shelf is up, and the cords are taut.

outfit was put together. The make-up is indicated in Fig. 1. The total mass of the yo-yo is 440 g. It consists of two brass cylinders about 5 cm in diameter, firmly set on a steel pin of diameter 4 or 5 mm. The pulley was of a common sort with a light fiber wheel on pin bearings, but it has been much improved in our shop by the insertion of a small ball bearing of good quality. With the yo-yo accurately counterbalanced, an extra load of about 5 g is required on either side to maintain uniform motion. The spring shelf, shown best in Fig. 1, serves to get the yo-yo and the weight off to an even start, a start essentially the same every time.

With this equipment the experiment frequently goes as Mackay indicates, and as Figs. 2, 3 and 4 show. That is, the yo-yo and the counterweight keep nearly in step. It does not perform well on every trial, but on the one occasion when I used it for a large class it worked perfectly. Sometimes the cord—a piece of fishline of good quality—jumps off the wheel.

Mackay does not indicate how good this experiment really is. It is a capital illustration of the fundamental law of mechanics, that the centroid of a body moves as if it contained the entire mass of the body and all the acting forces were applied, in their original magnitudes and directions, at the centroid. For if the pulley is ideal the cord exerts two equal, upward forces upon the yo-yo and the counterweight. If they balance when at rest the two weights are equal. Hence the resultant force on one is the same as the resultant force on the other. Therefore, since the two masses are equal, the two accelerations are always the same. Since the objects start together from rest, both their velocities and their displacements are always the same. The fact that the two forces on the counterpoise are in the same line, while those on the yo-yo are not, is irrelevant. The fact that the yo-yo rotates, while the counterpoise does not, has no influence on the motion of either centroid. The change in the effective diameter of the axle as the cord is reeled off or wound on changes the linear acceleration of each body equally.

Very probably there is a best value for the space between the two disks of the yo-yo, where the cord is wound, and perhaps there is a particularly good method of winding up

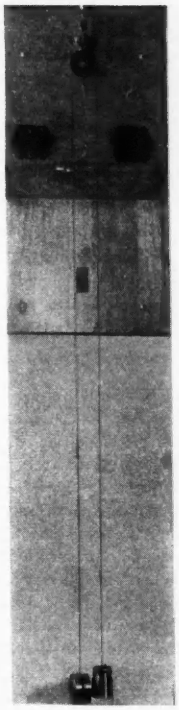

FIG. 3. Going down, still even near end of cord.

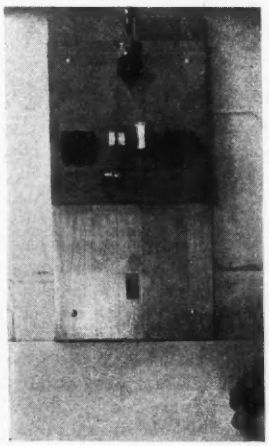

FIG. 2. Off to an even start.

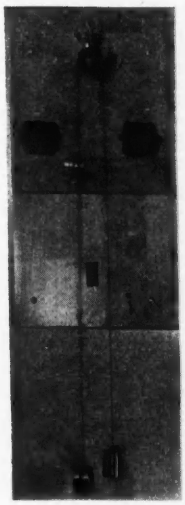

FIG. 4. Coming up, nearly even after two round trips.

the cord, but we have not yet had opportunity to check these and other promising variations. It is a good problem to give the mass and radius of the yo-yo and the effective radius of the axle and require the linear accelerations of the two centers. A more difficult but still better problem would be to estimate the effective diameter of the axle in terms of the length of cord unwound and then to calculate the time of descent.

[1] R. S. Mackay, *Am. J. Physics* **16**, 248 (1948).

Rocket Power

E. A. Cook, Jr.
Chemical Corps Technical Command, Army Chemical Center, Maryland

KENNARD[1] objects to attributing extremely high power to constant thrust motors (rockets) as they reach high velocities, when the frame of reference is considered attached to the launching point or other fixed point on the ground.

While it is true that power may be defined in fundamental mechanical terms only with respect to some arbitrary fixed reference system for velocity, this does not seem objectionable when the purpose of evaluating power is related to questions of ballistics: range, time of flight and so on, of a missile on or near the earth. Furthermore, it seems that any attempt to define power in terms of energy (chemical or thermal) and efficiency, must ultimately be rather circular; for how does one determine the efficiency except with the knowledge of the rate of development of useful work (in this case, kinetic energy)?

We take great pains to tell the elementary student that he does no mechanical work while pushing against a wall or holding up a weight, because the force he applies does not have motion associated with it. We point out that an engine is enabled to develop more power with the same torque by turning at a higher rate, with the aid of gears or an adjustable-pitch propeller. (This amounts to increasing the force available at a low velocity, thereby maintaining the product FV, a process not feasible with rockets.)

Must we now tell him that he does no work in supporting the weight because the efficiency of conversion of food energy into mechanical work is zero when he doesn't move, and that the burning of more fuel with the same efficiency in the engine is the measure of higher power? This seems simply to throw out the fundamental mechanical definition altogether. Why not admit that power is as relative to the frame of reference as are velocity and kinetic energy?

The purpose of stressing the high power of a rocket with high velocity, relative to the launching point, *and* the *low* power with *low* velocity, seems, I repeat, a legitimate one; since it highlights a fundamental characteristic of constant-thrust motors in use. They are inferior for take-off power, and superior for power at extremely high speeds—neglecting air drag, of course—when one is interested in the power for obtaining range, payload and speed of flight relative to the origin.

[1] R. B. Kennard, "Rocket power," *Am. J. Physics* **15**, 357 (1947).

Rocket Power and the Frame of Reference

Leonard T. Pockman
San Francisco State College, San Francisco, California

IN his note on rocket power, Kennard[1] calls attention to an apparent inconsistency in a statement by Seifert, Mills and Sommerfield.[2] These authors state:

> The propulsive power developed by a rocket is proportional to its speed. For example, a German V-2 rocket at its maximum speed of 5000 ft/sec develops over half a million horsepower, whereas the power immediately after take-off is relatively low.

Commenting on this, Kennard says:

> This statement assumes the equation: Power = Thrust × Velocity. This equation is valid only when there is a unique frame of reference to which the velocity must be referred. . . . There is no unique frame of reference outside of the rocket-gas system to which the velocity must be referred. Any frame of reference that may be chosen is arbitrary, and the velocity with respect to such a frame has nothing to do with the power output.

The purpose of this note is to show that there is no real contradiction between the constant power output of the rocket motor and the fact that the propulsive power of the rocket is strictly proportional to the velocity of the rocket with respect to any frame of reference moving with constant velocity.

Consider the kinetic energy of the rocket, neglecting relativity, and let v be the velocity with respect to a frame of reference moving with constant velocity relative to an inertial frame. For concreteness, assume that a frame of reference fixed in the earth approximates such an ideal frame sufficiently for the present purpose. If we denote the kinetic energy of the rocket by E_R, then

$$dE_R/dt = vm(dv/dt) + \tfrac{1}{2}v^2(dm/dt). \qquad (1)$$

The rate dm/dt at which the mass of the rocket changes is constant and is less than zero. Thus, the second term in Eq. (1) is negative and is proportional to v^2; and the first term is positive, since $dv/dt > 0$, and is proportional to v. This means that for a large enough value of v the kinetic energy of the rocket will be diminishing with time. This corresponds to the fact that propulsive power $vm(dv/dt)$ is not able to compensate for the loss of kinetic energy corresponding to the detachment from the rocket of mass moving with high velocity and therefore possessed of large kinetic energy. It further corresponds with the fact that for an ideal self-destroying rocket the kinetic energy becomes zero at a time $t = -m/(dm/dt)$. More specifically, the term $\frac{1}{2}v^2(dm/dt)$ is simply the kinetic energy of the mass disengaged per unit time from the rocket.

The velocity v_c at which dE_R/dt drops to zero on its way to becoming less than zero can be found by setting the derivative equal to zero; we obtain $v_c = (2mdv/dt)/(-dm/dt)$. If the escape velocity of the gases measured *relative to the rocket* be called v_e, Newton's third law leads to the equation $-v_e(dm/dt) = m(dv/dt)$. Hence, $v_c = 2v_e$.

Finally, it must be shown that the kinetic energy of the total system of rocket plus exhaust gases increases at a rate equal to the power of the rocket motor. The power of the motor is of course constant and is simply $\frac{1}{2}v_e^2(-dm/dt)$, as may easily be seen by imagining the rocket anchored in the rest system for an experimental test. If the kinetic energy of the exhaust gases be denoted by E_g, then

$$dE_g/dt = \tfrac{1}{2}(v-v_e)^2(-dm/dt),$$

since the gas separating at any instant from the rocket has a velocity relative to the earth of $v-v_e$, and the mass of the tail continually increases with time. Thus the rate of change of total kinetic energy is given by

$$\frac{dE_R}{dt}+\frac{dE_G}{dt}=-\tfrac{1}{2}v_e^2\frac{dm}{dt}+vm\frac{dv}{dt}+vv_e\frac{dm}{dt}$$

$$=-\tfrac{1}{2}v_e^2\frac{dm}{dt},$$

the power of the rocket motor.

[1] R. B. Kennard, *Am. J. Physics* **15**, 357 (1947).
[2] Seifert, Mills and Sommerfield, "Physics of rockets," *Am. J. Physics* **15**, 1 (1947).

Isochronous Pendulums: A Correction

W. W. Sleator
University of Michigan, Ann Arbor, Michigan

SINCE the appearance of my paper about pendulums,[1] my attention has been called to the article by Willoughby M. Cady,[2] entitled "A remarkable isochronous pendulum." Doctor Cady, in a treatment of isochronism much more general than mine, points out the fact, indicated in my Fig. 2, that the oscillation of the complete ring pendulum "is isochronous with respect to any alteration that leaves equal arcs on each side of the point of support." I am sorry that I did not refer to this article; it would have been done if I had known that the statement quoted was in it. On this point, and on isochronous variations of circular and elliptical disk pendulums, Doctor Cady makes valuable contributions.

[1] W. W. Sleator, *Am. J. Physics* **16**, 93 (1948).
[2] W. M. Cady, *Am. J. Physics* **10**, 114 (1942).

NECROLOGY

Homer W. LeSourd, 1875–1948

One of the greatest of secondary school teachers, HOMER W. LESOURD, died at his home in Cambridge, Massachusetts on March 15, 1948. For almost 50 years he was a leader among teachers in the New England Association of Physics Teachers, serving first as secretary and later as president. When a group of 22 physicists met at Cleveland on December 29, 1930 to organize the American Association of Physics Teachers, Mr. LeSourd was present and took part in laying the plans for the organization.

Born at Bellefontaine, Ohio, in 1875, educated in the public grammar and high schools of that city, and at Ohio Wesleyan University, he began his teaching career in Delaware, Ohio, in 1898. His M.A. degree was granted by Harvard University in 1901. After one year in Connecticut he accepted a place as science teacher in Milton Academy, Milton, Massachusetts, where he taught until the close of World War II, except for one year as an exchange teacher in the Punahou School of Honolulu. For a number of years he taught classes in the teaching of physics and of general science in the Summer School of Harvard University. At the close of the war he gave up his work at Milton but continued at Harvard, where he assisted in putting the optics laboratory in shape for better teaching and research.

In addition to being a charter member of the American Association of Physics Teachers, MR. LESOURD was a member of the Eastern Association of Physics Teachers and physics editor of *School Science and Mathematics*. His contributions to the leading educational and teaching journals of this country are numerous.

GLEN W. WARNER

RECENT MEETINGS

Indiana Section

The Indiana Section of the American Association of Physics Teachers met at DePauw University on April 24, 1948. Professor O. H. Smith presided during the presentation of the following program.

The program in applied science in the department of physics of Valparaiso University. A. R. THOMAS, *Valparaiso University.*

Curriculum problems of physics departments. JESSE C. HENDRICKS, *Franklin College.*

An undergraduate course in radiation physics. K. WATANABE, *Wabash College.*

Physics and physical science in a general education program for colleges. J. F. MACKELL, *Indiana State Teachers College.*

A laboratory experiment on the compound lens. M. E. HUFFORD, *Indiana University.*

A simple interference experiment. K. W. MEISSNER, *Purdue University.*

An elementary laboratory test of Thevenin's theorem. A. D. HUMMEL, *Ball State Teachers College.*

(a) A logical and unique way of presenting the problem of fitting eyeglasses. (b) A treatment of Doppler's principle. EARL MARTIN, *Hanover College.*

A suggestion on physics tests. JESSE C. HENDRICKS, *Franklin College.*

Cheating on examinations. C. H. HIRE, *Indiana University.*

Molecular structure and physical properties of plastics. DAVID TELFAIR, *Earlham College.*

The Indiana science talent search. R. W. LEFLER, *Purdue University.*

The average high school student's lack of mathematical training. J. B. HERSHMAN, *Valparaiso Technical Institute.*

Recent work of the committee on election procedures. JOSEPH D. ELDER, *Wabash College.*

A brief report of the Chicago meeting of the Association. O. H. SMITH, *DePauw University.*

The following 43 members and guests registered for the meeting.

J. W. Alinsky, Valparaiso Technical Institute; E. E. Bullis, Valparaiso Technical Institute; P. Braunschwerger, DePauw University; D. Compton, Wabash College; W. R. Davies, Lebanon High School; J. Davis, Wabash College; T. R. Davis, Indiana State Teachers College; L. C. De Vol, Valparaiso University; J. D. Elder, Wabash College; S. Elliott, Wabash College; J. M. Fulmer, DePauw University; L. B. Guernsey, Valparaiso Technical Institute; R. Hare, DePauw University; J. C. Hendricks, Franklin College; J. B. Hershman, Valparaiso Technical Institute; M. E. Hufford, Indiana University; A. D. Hummel Ball State Teachers College; R. Keeling, Wabash College; A. R. Knudsen, Valparaiso Technical Institute; R. W. Lefler, Purdue University; E. Martin, Hanover College; J. F. Mackell, Indiana State Teachers College; M. R. Murray, Ball State Teachers College; E. Nussbaum, Taylor University; C. D. Pamby, Valparaiso University; V. J. Raelson, Valparaiso University; D. Roller, Wabash College; J. Ross, University of Wisconsin; T. Sando, Wabash College; A. Schudert, Taylor University; W. Shepherdson, Huntington College; O. H. Smith, DePauw University; J. Spears, Wabash College; A. D. Sprague, DePauw University; K. Steele, Tri State Teachers College; D. Telfair, Earlham College; A. R. Thomas, Valparaiso University; G. D. Van Dyke, Earlham College; J. W. Voigt, Indiana State Teachers College; K. Watanabe, Wabash College; R. E. Winn, Indiana State Teachers College; L. N. Zechiel, DePauw University; S. J. Zissis, Purdue University.

At the business meeting, R. W. Lefler, of Purdue University, was elected representative of the Section on the Executive Committee of the Association. It was voted to hold the next meeting at Hanover College.

The meeting closed with a dinner held in Mason Hall.

M. E. HUFFORD
Secretary

Kentucky Section

The regular spring meeting of the Kentucky Association of Physics Teachers was held at the University of Louisville on April 15, 1948. Dr. Wm. S. Webb, University of Kentucky, presided. Thirty-two members and guests were in attendance. The following papers were presented:

A time standard for the physics laboratory. H. T. SMITH, *University of Louisville* and *Girdler Corporation.*

Maxwell's mesh method. P. C. OVERSTREET, *Morehead State College.*

Some remarks on the galvanometer. M. S. COHEN, *University of Kentucky.*

Some experiences in Government and industrial research. J. G. BLACK, *Eastern Kentucky State College.*

Simplifying filter circuits. H. J. DAILY, *Lafayette High School, Lexington.*

A seismograph, electromagnetic type. S. J. ALLEN, *University of Kentucky.*

A Wilson cloud chamber. R. W. WILLMOTT, *University of Kentucky.*

A small Van de Graaff generator. R. A. LORING, *University of Louisville.*

At the business meeting, the following officers were elected for 1948: *President*, J. G. BLACK, Eastern Kentucky State College; *Vice president*, R. A. LORING, University of Louisville; *Secretary-Treasurer*, L. W. COCHRAN, University of Kentucky; *Representative on Executive Committee*, D. M. Bennett, University of Louisville.

LEWIS W. COCHRAN
Secretary-Treasurer

District of Columbia and Environs

The ninth annual meeting of the District of Columbia and Environs Section of the American Association of Physics Teachers was held at the University of Maryland on April 10, 1948. The following papers were presented during the morning session.

A laboratory arts course in physics. RAYMOND MORGAN, *University of Maryland.*

Aerophysics. RAYMOND SEEGER, *Naval Ordnance Laboratory.*

How can interest in physics be aroused and maintained? R. R. MEIJER, *The George Washington University.*

A simple method of making ice in the laboratory. H. G. DORSEY, *C.R.E.I.*

The laboratory experiment as a "project." R. A. GOODWIN, *U. S. Naval Academy.*

Speaking of physics: simple demonstrations. G. M. KOEHL, *The George Washington University.*

Supersonic wind tunnels. G. L. SHUE, *Naval Ordnance Laboratory.*

The status of learning at the Naval Academy. EARL W. THOMSON, *U. S. Naval Academy.*

Proposed plans for physics instruction in District of Columbia public schools. KEITH JOHNSON, *District of Columbia Public Schools.*

Displacement currents. T. B. BROWN, *The George Washington University.*

A study in cosmic radiation. CHARLES KISSINGER, *University of Maryland.*

Of interest to the members attending the meeting was the inspection trip to the wind tunnel during the luncheon period.

The following officers were elected at the annual business meeting: *President*, G. M. KOEHL, *The George Washington University; Secretary*, E. R. PINKSTON, *U. S. Naval Academy; Executive committee*, VOLA BARTON, CLARE DRISCOLL, EARL W. THOMSON.

E. R. PINKSTON
Secretary

Oregon Section

The Oregon Section of the American Association of Physics Teachers held regular meetings on December 6, 1947 and February 28, 1948. Raymond T. Ellickson, President of the Section, presided.

At the December meeting, held at the University of Portland, 63 members and guests heard the following program.

Report on the conference of the Association of Land Grant Colleges and Universities. W. WENIGER, *Oregon State College.*

Measurement of index of refraction and thickness of optical films. FRED W. PAUL, *University of Oregon.*

Cosmic rays at high elevations. E. P. COOPER, *University of Oregon.*

Scientific training in the United States. RAYMOND T. ELLICKSON, *Reed College.*

"Atomic energy," a Navy training film. FRED W. DECKER, *Oregon State College.*

A magnetic lens beta-ray spectrograph of new design. E. G. EBBIGHAUSEN, *University of Oregon.*

Differentiating and integrating circuits. J. K. TROLAN, *Oregon State College.*

Calorimetric measurement of beta- and gamma-energies. C. V. CANNON and G. H. JENKS, *Clinton Laboratories.*

At the February meeting, held at Linfield College, 51 persons were present. The program follows.

The Argonne Laboratory. R. L. PURBRICK, *Willamette University.*

Pictures of the V-2 rocket. MARCUS O'DAY.

Teaching physics for industry. E. H. COLLINS, *Longview, Washington.*

Some problems involving electrons in metals. J. J. BRADY, *Oregon State College.*

University of Tucuman in Argentina. G. VASSALO, *University of Portland.*

Field emission. W. P. DYKE, *Linfield College.*

Progress report on graduate fellowships in the atomic energy field. W. WENIGER, *Oregon State College.*

Report on the annual meeting of the American Association for the Advancement of Science. W. V. NORRIS, *University of Oregon.*

During the business meeting the invitation of the Department of Physics of the University of Washington to meet there in May 1948 was accepted.

W. R. VARNER
Secretary

Southern California Section

The regular spring meeting of the Southern California Section of the American Association of Physics Teachers was held in Science Hall at the University of Redlands on March 13, 1948. Sixty-six persons attended the meeting. Fifty members and guests attended the luncheon given at the University between the morning and afternoon sessions.

The following program of 10-min contributed papers was heard in the morning session.

The passage of time during man's life span. F. R. HIRSH, JR. *Pasadena.*

Student opinion as used at University of Southern California to improve physics laboratory instruction. W. GEER, *University of Southern California.*

Education—or merely training. Part IV: Newton's second law. G. FORSTER, *Pasadena City College.*

The break-up of water jets. J. M. SCHMIDT, *Jet Propulsion Laboratory, California Institute of Technology.*

The morning session closed with an invited paper by H. P. Robertson, Professor of Mathematical Physics at the California Institute of Technology, who spoke on "Geometry and physical space" at a joint meeting with the Southern California Branch of the Mathematical Association of America.

The following program of contributed papers was heard in the afternoon session.

Elementary derivation of the uniaxial and biaxial wave surfaces. S. BROWN, *University of California at Los Angeles.*

Development of a lecture demonstration experiment. L. E. DODD, *University of California at Los Angeles.*

An experiment on 60-c/sec resonance for general and engineering physics.—W. GEER, *University of Southern California.*

Report on the Annual Meeting of the Association at Chicago. D. L. SOLTAU, *University of Redlands.*

At the business meeting following the afternoon session, V. L. Bollman, Occidental College, reported on the work of the High School Physics Test Committee. The test has been scheduled for Saturday, May 15, 1948 from 9:00 to 11:00 A.M. at University of California at Los Angeles, University of California at Santa Barbara, University of Redlands, and San Diego State College. Scholarships have been offered to the winners of this Fourth Annual High School Physics Test by California Institute of Technology, University of California (any campus), University of Southern California, Occidental College, Pomona College, Redlands University, and Whittier College. There will be a regular program meeting for high school instructors in connection with the test at U.C.L.A., but no instructor meetings were planned for the other schools.

Earle C. Enholm, of Long Beach City College, reported for the Teaching Load Committee. He told of the large amount of work that had been done in framing the form of the questionary to be sent to all high school and junior college instructors of physics in the Southern California area. Mr. Enholm announced that this questionary would also be sent later to a sample group of college instructors, to determine their definition of a proper teaching load.

The following officers were elected to serve for the year 1948–49: *President*, R. W. McHenry, *Santa Monica City College; Vice President for Colleges*, A. W. Nye, *University of Southern California; Vice President for Junior Colleges*, G. O. Sagen, *Bakersfield Junior College; Vice President for High Schools*, J. G. Betts, *J. C. Fremont High School, Los Angeles; Secretary-Treasurer*, F. Strong, *California Institute of Technology.*

FOSTER STRONG
Secretary-Treasurer

DIGEST OF PERIODICAL LITERATURE

The Physicist and the Future Development of Atomic Energy

From 1940 until late in 1945, the traditions of free and open research in the sciences were laid aside in favor of goals useful in the war. At the end of the war, many scientists tired of working at classified projects were eager to get back to the prewar, unrestricted scientific atmosphere and felt that many of the subjects considered classified during the war would no longer need to be kept as such.

In the Atomic Energy Project, a committee, headed by Dr. Richard Tolman, was set up to recommend policies of declassification. First meeting in the fall of 1945, the committee recommended certain sections of basic physics, chemistry, metallurgy, biology and certain technological developments for declassification. It felt that technical information connected with the development of atomic weapons could not be declassified until a satisfactory international arrangement for the control of atomic energy had been developed. However, as the difference of opinion among nations regarding such control is deep and fundamental, a policy adopted now for unrestricted research and for development in the classified fields should be more than an interim policy.

Since research and development in many diverse subjects are vital to progress in the Atomic Energy Project, much thought in the United States Atomic Energy Commission has been directed toward the relation between classified and unclassified work. The duties of the Commission as they appear today can be divided into the following categories:

1. *The production of fissionable materials.*—This depends upon the maximum utilization of the raw material and upon the development of new sources of raw materials.

2. *The development, production, and testing of atomic weapons.*—Development and production of atomic weapons is carried on at the Los Alamos Laboratory in New Mexico and testing at the recently established proving ground at Eniwetok in the Pacific.

3. *Development of nuclear reactors.*—Nuclear reactors, the producers of fissionable material, are also the producers of heat and perhaps of electrical power on a large-scale, practical basis.

4. *Basic research.*—A sound, fundamental understanding of many sciences is necessary for any long-range technical development.

One of the decisions of the Commission aimed at carrying out this program effectively is to establish a considerable number of fellowships in the physical, biological and medical sciences, as well as in the new field of health physics. These are to be administered through the National Research Council.

Basic research in sciences related to atomic energy should be centered in the universities and the unrestricted areas of the Commission-owned Laboratories. Classified research, on the other hand, should be confined as far as possible to the laboratories established for that purpose, and for security reasons at relatively few central locations. Therefore, the development of weapons and reactors and the production of fissionable materials will be carried out most likely in the installations at Los Alamos, Oak Ridge, Hanford, and the Argonne Laboratory.

During the past year several requests have been made to the Commission for an exact demarcation between classified and unclassified information. The Commission does not believe that all nuclear physics is or should be classified. Although it believes that work in most areas not touching on fissionable materials or other special products of the Atomic Energy Project will avoid classified information, the Commission cannot guarantee that all results will remain in the unrestricted class.

The Commission will not grant restricted data to any individual until, as stated in the McMahon Act ". . . the Federal Bureau of Investigation shall have made an investigation and report to the Commission on the character, associations, and loyalty of such individual and the Commission shall have determined that permitting such person to have access to restricted data will not endanger the common defense or security."

For scientists in particular, the question of clearance raises definite problems, and in order to avoid injustices, the Commission has invited a group of distinguished citizens, under the chairmanship of Owen J. Roberts, to act as a Board of Review. Other members of this Board are Karl T. Compton, Joseph C. Grew, George M. Humphrey, and H. W. Prentis, Jr.

As for the work of the Commission during the past year, the downward trend which began at the end of the war seems definitely reversed. At Los Alamos there has been significant new work on the development of weapons and work in connection with the new proving ground in the Pacific. At Hanford, there is in progress a major construction program, replacing obsolete units and taking advantage of new developments. There have also been many new contributions in research.

We see that some of the efforts to change this gigantic project to a peacetime development are beginning to show results. We have called upon many for help during this past year and the unselfish response has been very heartening. The General Advisory Committee, in particular, has been invaluable in this work. We believe that the rate of advance increased during 1947 and are confident that it will be still further increased during the coming year.—*Summary of a speech by* ROBERT F. BACHER *before the American Physical Society, New York, January 30, 1948.*—H. R. W.

Atomic Energy Commission

A lively account of the activities of the Commission is given in the *New Yorker* (March 20, 1948), p. 58.